

Percepties van Formatief Assessment:
Constructie en Validatie van een Vragenlijst

Perceptions of Formative Assessment:
Construction and Validation of a Questionnaire

Sandra Broers

1 Maart, 2015

Master Onderwijswetenschappen

Open Universiteit Nederland

Begeleiders: Dr. D. Joosten-ten Brinke

Dr. J. Castelijns

Examinatoren: prof. dr. H.P.A. Boshuizen

prof. dr. S. Brand-Gruwel

Studentnummer: 851335033

'Door het inleveren van dit voorstel verklaar ik dat het eigen werk is en dat het vrij is van plagiaat.'

‘Disciplina vitae scipio’ (Onderwijs is een staf voor het leven)

Samenvatting

Hogeschool de Kempel, een lerarenopleiding basisonderwijs, maakt een omslag van een testcultuur naar een assessmentcultuur, waarbij de nadruk ligt op integratie van toetsen en beoordelen met instructie en leren door actieve, betrokken lerenden. Eén van de factoren die van invloed is op de betrokkenheid bij deze integratie is de perceptie van studenten en lerarenopleiders van methoden, manieren en hoeveelheid van assessment. Kennis van studentpercepties kan bijdragen aan bewustwording bij lerarenopleiders van eigen gedrag en aanpassing van leren en onderwijs. De verwachting is dat door deze aanpassing het leerproces bij studenten verbeterd kan worden en daarmee de kwaliteit van leren en beoordelen in het onderwijs. Er is echter minimaal onderzoek gedaan naar studentpercepties en daardoor zijn er geen valide meetinstrumenten beschikbaar om studentpercepties van formatief assessment te meten.

Het doel van dit onderzoek is het construeren en valideren van een vragenlijst bestaande uit een samenhangende set van schalen die inzicht geeft in studentpercepties van de mate waarin formatieve assessment wordt toegepast door lerarenopleiders. Hiermee wordt beoogd dat het instrument een hulpmiddel kan zijn om het leerproces bij studenten te verbeteren en om lerarenopleiders te professionaliseren.

De eerste stap bestond uit het selecteren en definiëren van componenten van formatief assessment, gebaseerd op de modellen van Hattie en Timperley (2007) en Black en Wiliam (2009) die de theoretische basis vormden voor het construeren van een initiële verzameling van items. In de tweede stap werden deze items voorgelegd aan lerarenopleiders en experts, inhoudsdeskundigen op het gebied van formatief assessment, met de vraag of in hun ogen de items een goede operationalisering zijn van de verschillende componenten. De derde stap was de afname van de vragenlijst 'Percepties van Feedback' bij de reguliere voltijdstudenten van leerjaar 1 tot en met leerjaar 4 van Hogeschool de Kempel. De vierde stap bestond uit statistische analyse van de data.

De respondenten werden willekeurig ingedeeld in twee groepen om bij de eerste groep (N = 278) de onderliggende structuur van de items te onderzoeken met een Exploratieve Factor Analyse (EFA) en om bij de tweede groep (N = 277) de stabiliteit van de gevonden factor structuur te beoordelen en valideren met een Confirmatieve Factor Analyse (CFA). Vervolgens werden de interne consistenties bepaald en inzicht in studentpercepties verworven.

Het resultaat is een vragenlijst met 29 items en een zes-factorenstructuur bestaande uit lerarenopleider-, peer- en self-feedback op taak- en procesniveau en op zelfregulatie-niveau. De gemiddelde scores van alle klassen op de subschalen feedback op zelfregulatie-niveau bleken onderscheidend van

en lager dan de subschalen feedback op taak- en procesniveau. Deze verschillen significant tussen alle leerjaren op de subschalen op zelfregulatie-niveau en tussen de leerjaren 1 en 4 op de subschalen op taak- en procesniveau. De gemiddelden binnen elk leerjaar verschillen significant, behalve van de klassen binnen het derde leerjaar.

De vragenlijst kan een bruikbaar instrument zijn voor lerarenopleiders om een beeld te krijgen van de mate waarin studenten waargenomen hebben dat ze feedback op taak-, proces- of zelfregulatie-niveau hebben ontvangen van een lerarenopleider, medestudent of henzelf.

Keywords: Percepties, Formatief assessment, Assessment for Learning, Hoger Onderwijs,
Factor analyses, Validatie

Perceptions of Formative Assessment: Construction and Validation of a Questionnaire

Sandra Broers

Abstract

Hogeschool de Kempel, an institute for primary teacher education, makes a shift of a test culture to an assessment culture, with an emphasis on integration of testing and assessment with instruction and learning by active, involved learners. One of the factors that affect the involvement in this integration is the perception of students and teacher educators of methods and assessment. Knowledge of students' perceptions may contribute to awareness of behaviour from the teacher educator and adaptation of learning and education. The expectation is that the learning process can be improved through this adaptation and thus the quality of learning and assessment. However, there is minimal research into students' perceptions and therefore there are no valid instruments available to measure students' perceptions of formative assessment.

The aim of this research is to construct and validate a questionnaire consisting of a coherent set of scales that provide insight into students' perceptions of the extent to which formative assessment is used by teacher educators. The expectation is that the instrument can be a tool to enhance learning among students and to professionalise teacher educators.

The first step consisted of selecting and defining component of formative assessment, based on the models of Hattie and Timperley (2007) and Black and Wiliam (2009) which formed a theoretical basis for the construction of an initial set of items. In the second step, these items submitted to teacher educators and experts, content experts in the field of formative assessment, with the question whether, in their view, the items are a good operationalization of the various components. In the third step the 65-item questionnaire 'Percepties van Feedback' ('Perceptions of Feedback') was presented to the regular full-time students from grade 1 to grade 4 of Hogeschool de Kempel. The fourth step consisted of statistical analysis of the data.

The respondents were randomly divided into two groups. The first group (N = 278) was used to explore the underlying structure of the items with an Exploratory Factor Analysis (EFA) and data from the second group (N = 277) were used to test if each factor was confirmed by Confirmatory factor Analysis (CFA). Followed by the determination of the internal consistencies and understanding of the students' perceptions.

The result is a questionnaire of 29 items with a six-factor structure consisting of teacher educator, peer- and self-feedback on task and process level and on self-regulation level. The average scores of all classes on the subscales feedback on self-regulation level were distinctive and lower than the subscales feedback on task and process level. These differ significantly between all years on the subscales

of self-regulation level and between grades 1 and 4 on the subscales on task and process level. The averages within each grade differ significantly, except for the classes within the third grade.

The questionnaire can be a useful tool for teacher educators to gain insight of the extent to which students have experienced that they received feedback on task and process level or self-regulation level of a teacher educator, peer or themselves.

Keywords: Perceptions, Formative assessment, Assessment for Learning, Higher education, Factor Analyses, Validation

Inhoudsopgave

Percepties van Formatief Assessment: Constructie en Validatie van een Instrument.....	1
Percepties	3
Theoretisch model van Formatief Assessment.....	5
Doelstelling	9
Vraagstellingen.....	9
Eisen aan vragenlijst	10
Percepties van feedback	11
Methode.....	12
Participanten.....	12
Procedure.....	13
Ontwikkeling vragenlijst.....	13
Stap 1: ontwikkeling items.....	13
Stap 2: review items	13
Stap 3: afname vragenlijst.....	14
Stap 4: statistische analyse	15
Data-analyse	15
Vorbereidende analyse	15
Exploratieve factoranalyse	15
Confirmatorische factoranalyse	16
Betrouwbaarheidsanalyse en beschrijvende statistiek.....	16
Correlaties en ANOVA	16
Resultaten.....	16
Deelvraag 1: Betrouwbaarheid en Validiteit.....	16
Vorbereidende analyse	17
Exploratieve factoranalyse	17
EFA vragenlijst lerarenopleider	17
EFA vragenlijst peer-assessment.....	18
EFA vragenlijst self-assessment.....	20
Confirmatorische factoranalyse	21
CFA vragenlijst lerarenopleider	21
CFA vragenlijst peer-assessment	22
CFA vragenlijst self-assessment	23

Betrouwbaarheidsanalyse.....	24
Deelvraag 2: Mate van Studentpercepties	25
Deelvraag 3: Samenhang.....	27
Deelvraag 4: Verschillen in Studentpercepties tussen Leerjaren	27
Deelvraag 5: Verschillen in Studentpercepties tussen Klassen binnen Leerjaar	28
Conclusie en Discussie.....	30
Beperkingen en Aanbevelingen.....	34
Referenties	37
Bijlagen.....	42
Bijlage A: Verzoek tot deelname aan onderzoek	42
Bijlage B: Eerste versie van de vragenlijst.....	43
Bijlage C: Items van de vragenlijst ‘Percepties van Feedback’	45
Bijlage D: Toelichting voor tutoren op het doel en de werkwijze van het onderzoek	48
Bijlage E: Vragenlijst ‘Percepties van Feedback’	49
Bijlage F: Items van de zes gevalideerde subschalen.....	50

Kwaliteit van beoordelen en onderwijs is en blijft een actueel onderwerp van gesprek in de onderwijspraktijk en in de huidige maatschappij. Met de actieplannen voor het basisonderwijs ‘Basis voor Presteren’, voor het voortgezet onderwijs ‘Beter Presteren’ en voor leerkrachten van alle onderwijssectoren ‘Leraar 2020 – Een krachtig beroep!’ is ingezet op opbrengstgericht werken, de Nederlandse term voor *Data-Driven Decision Making* (Ministerie van OC&W, 2011; Vermeulen & Van der Kleij, 2012). Opbrengstgericht werken betekent op mesoniveau dat scholen en leerkrachten regelmatig de vorderingen van hun leerlingen bepalen en de uitkomsten daarvan gebruiken bij de inrichting van verdere onderwijsactiviteiten (Blok, Ledoux, & Roeleveld, 2013). Een belangrijk aspect hiervan op microniveau is formatief assessment; hiervan is sprake als leerlingen bewust worden gemaakt van de mate waarin hun leerproces bijdraagt aan het bereiken van gestelde leerdoelen. Dit gebeurt aan de hand van verzamelde informatie via feedback van de leerkracht, peer-assessment of self-assessment (Andersen, Castelijns, & Van Ravenswaaij, 2013).

Ondanks de vele beschikbare studies, ervaringen en praktische handboeken waarin wordt aangetoond dat formatief assessment een effectieve leerstrategie is, is formatief assessment nog niet structureel ingebed in de Nederlandse onderwijscurricula of onderdeel van politieke en maatschappelijke beleidsontwikkelingen (Robinson, Myran, Strauss, & Reed, 2014; Sluijsmans, Joosten-ten Brinke, & Van der Vleuten, 2013). De toepassing van formatief assessment in scholen vereist namelijk een conceptuele *mindshift* bij alle betrokkenen, van een focus op meten van leerprestaties in de vorm van summatief toetsen naar een focus op het proces van leren, waarbij de nadruk ligt op het verbeteren van het leerproces van leerlingen met een gezamenlijke verantwoordelijkheid van leerlingen, leerkrachten en medeleerlingen (Birenbaum et al., 2006; Black & Wiliam, 2009; Rusman, Boon, Martínez-Monés, Rodríguez-Triana, & Retalis, 2013; Segers & Tillema, 2011). Bovendien blijken leerkrachten over het algemeen niet of nauwelijks geschoold te zijn in formatief assessment, waardoor de kracht van formatief assessment nog onvoldoende wordt benut in de praktijk (Bennett, 2011; Birenbaum, Kimron, & Shilton, 2011).

Lerarenopleidingen zijn bezig met de vraag op welke manier ze bij studenten de kennis van formatief assessment kunnen vergroten. Hogeschool de Kempel, een lerarenopleiding basisonderwijs, heeft de module ‘Beoordelen om te leren’ ontwikkeld, die structureel aan alle studenten in het tweede leerjaar wordt aangeboden (Broers, 2014; Castelijns, Heldens, Wijnen, & Wouters, 2011). Studenten leren een eigen visie op toetsen en beoordelen te formuleren, waarbij het onderscheid tussen de summatieve functie en de formatieve functie van beoordelen duidelijk is. Ze ontwikkelen leerkrachtvaardigheden en elementen van instructie die nodig zijn om beoordelen om te leren te stimuleren bij kinderen.

Daarnaast vragen deze lerarenopleidingen zich af hoe een omslag gemaakt kan worden van een testcultuur naar een assessmentcultuur, waarbij de nadruk ligt op integratie van toetsen en beoordelen

met instructie en leren door actieve, betrokken lerenden. Dat heeft tot gevolg dat zij onderzoeken hoe formatief assessment bij kan dragen aan kwaliteitsverhoging van het onderwijs op de lerarenopleiding zelf en dat zij zoeken naar manieren om de betrokkenheid van zowel studenten als lerarenopleiders bij de integratie van leren en beoordelen te versterken.

Eén van de factoren die van invloed is op de betrokkenheid bij deze integratie is de perceptie van studenten en lerarenopleiders van methoden, manieren en hoeveelheid van assessment, en niet zozeer van de context waarin leren en beoordelen plaats vindt (Entwistle, 1991; Lizzio & Wilson, 2004). De perceptie bepaalt in welke mate er geleerd wordt door zowel studenten als lerarenopleiders. Dus om te begrijpen of en hoe formatief assessment werkt in de praktijk is het noodzakelijk om inzicht te krijgen in de percepties van studenten (Haroldson, 2012). Als lerarenopleiders begrijpen hoe formatief assessment leren verbetert en motiveert en hoe studenten het ervaren, kunnen ze effectief en succesvol strategieën aanbieden (Haroldson, 2012; Hattie, 2013; Pat-El, Tillema, Segers, & Vedder, 2013).

Formatief assessment is een continu proces van interactie tussen studenten en lerarenopleiders, waarin feedback geven, ontvangen en gebruiken een belangrijke rol speelt (Black & Wiliam, 2009). De boodschap die in het feedbackproces wordt gegeven wordt ook gefilterd door de perceptie van de student en lerarenopleider, beïnvloed door voorkennis, ervaring, motivatie, om er vervolgens betekenis aan te geven. Uit verschillende onderzoeken is gebleken dat er verschil is in perceptie tussen studenten en lerarenopleiders van de mate waarin assessment is geïntegreerd in instructie en lesinhoud (Könings, Brand-Gruwel, & Van Merriënboer, 2005; Schulz, 2001). Een recent onderzoek van Korver en Tillema (2014) bevestigt deze discrepantie; de lerarenopleiders beoordelen de mate van feedback hoger dan de studenten. Als lerarenopleiders overtuigd zijn dat zij hun studenten constructieve feedback hebben gegeven en duidelijke doelen hebben gecommuniceerd, maar de studenten percipiëren dit niet, dan is het effect van hun begeleiding en het leren van de student verminderd (Pat-El et al., 2013). Als er geen congruentie is tussen percepties van studenten en lerarenopleiders leidt dat tot verkeerd begrijpen, misinterpretaties van de assessmentinformatie, de betekenis en het doel (Loughran, 2012). Kennis van studentpercepties kan dus bijdragen aan bewustwording bij lerarenopleiders van eigen gedrag en aanpassing van begeleiding en instructie. De verwachting is dat door deze aanpassing het leerproces bij studenten verbeterd kan worden en daarmee de kwaliteit van leren en beoordelen in het onderwijs verhoogd kan worden (Struyven, Dochy, & Janssens, 2005).

Uit voorgaande kan geconcludeerd worden dat het belangrijk is om percepties van studenten te meten om het effect van formatief assessment te vergroten. Er is echter minimaal onderzoek gedaan naar studentpercepties en daardoor zijn er geen valide meetinstrumenten beschikbaar om studentpercepties van formatief assessment te meten (Perumanathan, 2014). In dit onderzoek staat het construeren en valideren van een vragenlijst centraal dat studentpercepties meet van de mate waarin formatief assessment wordt toegepast op een lerarenopleiding basisonderwijs. De begrippen met bijbehorende ken-

merken die in deze vragenlijst gebruikt worden en eisen waaraan deze vragenlijst moet voldoen, worden hierna in het theoretisch kader beschreven.

Percepties

In de review van Struyven, Dochy en Janssens (2003) wordt aangegeven dat in verschillende studies verschillen zijn aangetroffen tussen operationalisering en definiëring van het begrip ‘perceptie’ van assessments met zowel een summatieve als formatieve functie. Sommige onderzoekers definiëren perceptie als de mening van leerlingen en leerkrachten over leren en toetsen door bijvoorbeeld te vragen of ze formatief assessment gerechtvaardigd vinden. Andere onderzoekers benadrukken de houding ten opzichte van en de voorkeuren voor bepaalde vormen van leren en toetsen door bijvoorbeeld te vragen naar de voorkeur voor eind- of tussentoetsen. Deze definities sluiten echter meer aan bij het begrip ‘conceptie’, dat opvattingen of attitudes omvat over bijvoorbeeld formatief assessment.

In dit onderzoek wordt perceptie gedefinieerd als waarneming. Het is het vermogen om te zien, te horen of je bewust te worden van iets door de zintuigen en de manier waarop het wordt begrepen en geïnterpreteerd (oxforddictionaries, 2015). Dus perceptie is het proces waarbij stimuli worden geselecteerd, georganiseerd en geïnterpreteerd om er een betekenisvol geheel van te maken dat iedere lerende op zijn eigen, unieke manier doet (Valcke, 2010). Het sluit aan bij de constructivistische theorie waar kennis wordt gezien als een proces waarbij elke lerende betekenis probeert te geven aan zijn of haar ervaringen en dit op een persoonlijke manier verwerkt, afhankelijk van de context. Het gaat om het subjectieve, individuele beeld van de werkelijkheid om iemand heen en dat beeld kan op basis van ervaringen of nieuwe kennis veranderen. De perceptie is daarmee afhankelijk van de interactie tussen een individueel persoon met zijn of haar omgeving, dus datgene wat iemand waar neemt in combinatie met zijn of haar omgeving (Driscoll, 2005). In dit onderzoek staat de interactie tussen lerenden, de studenten en de lerarenopleiders, centraal.

Het is aangetoond dat juist de perceptie van de leeromgeving, hoe deze door leerlingen en leerkrachten ervaren wordt, en niet de concrete leeromgeving, effect heeft op leerprocessen van leerlingen (Entwistle, 1991). Hierbij zijn niet zozeer de kenmerken van de leeromgeving op zich bepalend voor het studeergedrag, zoals didactische werkvormen en verwachte evaluatiemethoden, maar wel de manier waarop leerlingen deze werkvormen en evaluatievormen percipiëren en interpreteren. Hoe leerlingen deze percipiëren hangt af van verschillende factoren als kenmerken, doelstellingen of criteria van het assessment, kenmerken van de leerling zelf, zoals motivatie, leerstijl, intelligentie of voorkennis, en de eigenschappen van de onderwijsleeromgeving zoals de gebruikte methode, de beschikbare middelen of kenmerken van de leerkracht (Struyven et al., 2003). Hoewel een leeromgeving krachtig ontworpen en uitgevoerd kan zijn, zal de perceptie van de leerlingen van de leeromgeving bepalen wat de kwaliteit van de leerresultaten zal zijn (Könings et al., 2005). De perceptie kan zich negatief en positief evolueren, inzicht hierin is effectief voor het verbeteren van het leren en beoordelen in het

onderwijs (Joseph, Yakhou, & Stone, 2005). Als de kwaliteit van formatief assessment in de perceptie van leerlingen en leerkrachten voldoende vervuld is, is er tevredenheid over formatief assessment (Matzler & Hinterhuber, 1998). Vanuit het gezichtspunt van leerlingen heeft assessment een positief effect op hun leren als het om authentieke taken gaat, het redelijke eisen representeert, hen stimuleert kennis toe te passen op realistische contexten, de nadruk ligt op de noodzaak om een brede range van vaardigheden te ontwikkelen en het op lange termijn voordelen oplevert (Struyven et al., 2003). Voorgaande rechtvaardigt waarom het belangrijk is om percepties te meten.

Pat-El et al. (2013) hebben al wel vragenlijsten ontwikkeld om de percepties van formatief assessment bij leerlingen en leerkrachten te meten. Reden hiervoor was dat uit eerdere onderzoeken bleek dat beschikbare vragenlijsten als de *Assessment Experience Questionnaire* van Gibbs en Simpson (2003) en de vragenlijst *Learning how to Learn* van James en Pedder (2006) onvoldoende hebben aangetoond welke principes van formatief assessment worden gepercipieerd door leerlingen en leerkrachten. (Pat-El et al., 2013). Daarnaast was gebleken dat tussen leerlingen en leerkrachten verschil bestaat in perceptie van de mate waarin assessment is geïntegreerd in instructie en lesinhoud. Er is dus vaak een discrepantie tussen het onderwijs zoals het bedoeld was en het onderwijs zoals uitgevoerd door de docent of als beleefd door de betrokken leerlingen (Van den Berg & Vandenbergh, 1999). Maar er bleek een gebrek aan kwantitatieve instrumenten te zijn om de percepties van leerlingen te vergelijken met de percepties van leerkrachten (Könings et al., 2005; Schulz, 2001). Uit later onderzoek van Korver en Tillema (2014) blijkt ook een discrepantie in percepties tussen leerlingen en leerkrachten, de leerkrachten beoordelen de mate van feedback hoger dan de leerlingen.

De ontwikkelde vragenlijsten, de *Teacher Assessment for Learning Questionnaire* (TAFL-Q) en de *Student Assessment for Learning Questionnaire* (SAFL-Q), meten de houding van leerkrachten en leerlingen ten aanzien van twee aspecten van formatief assessment, namelijk *monitoring* en *scaffolding*, gebaseerd op enkele publicaties (ARG, 2002; CCSO; Darling-Hammond, 2010; OECD, 2005; Prosser & Trigwell, 1993 in Pat-El et al., 2013). Bij *scaffolding* wordt feedback aan de leerling teruggekoppeld, zodat deze zelf inzicht verwerft in wat hij goed kan en waar hij nog aan kan werken. Een voorbeeld van een item over *scaffolding* is: 'Ik zorg ervoor dat de leerling weet aan welke punten hij of zij moet werken om zijn of haar resultaten te verbeteren'. Bij *monitoring* vindt er geen terugkoppeling plaats en beslist de leerkracht over te nemen vervolgstappen. Een voorbeeld van een item over *monitoring* is: 'Ik bespreek met mijn leerlingen hun vorderingen'. De TAFL-Q en de SAFL-Q bleken in het onderzoek van Pat-El et al. (2013) en Korver en Tillema (2014) valide vragenlijsten om de twee componenten van formatief assessment te meten.

Naar aanleiding van deze onderzoeken kunnen enkele kanttekeningen geplaatst worden waaruit vragen voortvloeien die aanleiding zijn voor onderhavig onderzoek. Pat-El et al. (2013) geven aan dat er meerdere studies nodig zijn om de huidige dimensionaliteit van formatief assessment te valideren.

De twee factoren laten geen compleet beeld zien van de bouwstenen van formatief assessment. Het is de vraag of de gekozen dimensies van formatief assessment (*monitoring* en *scaffolding*) voldoende laten zien hoe leerlingen en leerkrachten formatief assessment percipiëren, met de wetenschap dat de definitie van formatief assessment nog steeds in ontwikkeling is en er meerdere interpretaties aan worden gegeven (Johnson & Burdett, 2010; Perumanathan, 2014). Ten tweede is de steekproef die zij gehanteerd hebben om hun vragenlijst te valideren, het voorbereidend middelbaar onderwijs, niet generaliseerbaar naar het hoger onderwijs. De vraag is of de vragenlijsten in het hoger onderwijs gevalideerd kunnen worden. Ten derde bleek de vragenlijst contextgevoelig te zijn, omdat de vragen betrekking hadden op actuele assessment activiteiten direct gerelateerd aan het leerproces in de klas. De percepties van leerlingen zijn genest in klaslokalen en antwoorden op de vragen zijn afhankelijk van bijvoorbeeld de kwaliteit van leerkrachten of het bewustzijn van doelen bij de leerlingen. De vraag is of de vragenlijst de verschillen in perceptie van formatief assessment kan onderscheiden.

Theoretisch Model van Formatief assessment

In de huidige literatuur en onderwijspraktijk worden in het kader van leren en beoordelen verschillende termen gehanteerd, namelijk ‘formatief toetsen’, ‘formatief beoordelen’, ‘formatief evalueren’ en ‘formatief assessment’, waarbij dubbelzinnigheid en verschil in connotatie ontstaat. In dit onderzoek is voor de term ‘formatief assessment’ gekozen, omdat deze term beter aansluit bij de *mindshift* van testnaar assessmentcultuur en gebaseerd is op de volgende definities van de termen ‘toetsen’, ‘beoordelen’, ‘evalueren’ en ‘assessment’. ‘Toetsen’ is een keuze maken voor een instrument om te meten, ‘beoordelen’ is een waardering aan een meting toekennen en ‘evalueren’ is het waarnemen van feiten en hun verbanden om een interpretatie te kunnen geven (Van der Kleij, Vermeulen, Eggen, & Veldkamp, 2013). Assessment gaat een stap verder, naast het beoordelen en evalueren worden leren en instructie geïntegreerd, het is dus een bredere vorm van evalueren. Voorbeelden hiervan zijn portfolio en peer-assessment waarbij het leren van studenten wordt bevorderd en verdiept (Struyven, Dochy, Janssens, Schelfhout, & Gielen, 2006). De term ‘assessment’ sluit meer aan bij het meten van leerprocessen en is een proces van kennisconstructie in plaats van kennisreproductie. Deze term is ontleend aan het Latijnse werkwoord *ad sedere* wat ‘er naast zitten, op gelijke hoogte zitten, aandacht besteden aan’ betekent. Het impliceert het verzamelen en interpreteren van informatie, het geven van begeleiding en feedback aan de lerende, met als functie het verder brengen van de lerende (Dochy, Schelfhout, & Janssens, 2003).

Ook worden in de literatuur verschillende definities van de term ‘formatief assessment’ gehanteerd, wat heeft geleid tot misvattingen in onderwijsbeleid en praktijk Klenowski (2009). Daarom is volgens Bennett (2011) formatief assessment conceptueel en praktisch een ‘werk in uitvoering’ en is een onderbouwde definitie van belang. Voor een betekenisvolle definitie is ten eerste een actietheorie nodig die kenmerken en componenten van formatief assessment identificeert. Toegepast op dit onderzoek

zijn de belangrijkste kenmerken en componenten de actoren lerarenopleiders, studenten en medestudenten (peers), hun onderlinge communicatie en het verbeteren van het leerproces. Ten tweede is een concrete uitwerking nodig hoe deze kenmerken en componenten samen een gewenst resultaat op kunnen leveren. Dus toegepast op dit onderzoek hoe formatief assessment op basis van de actietheorie eruit ziet en werkt in de onderwijspraktijk. De gekozen actietheorie en concrete uitwerking in dit onderzoek wordt hierna verder uitgewerkt.

In dit onderzoek vormt de definitie van formatief assessment geformuleerd door Black en Wiliam (2009) het uitgangspunt. Zij definiëren formatief assessment als volgt: *'Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited'* (Black and Wiliam, 2009, p.9). In deze definitie worden alle vormen van formatieve praktijken benoemd, maar daarbij wordt nog geen onderscheid gemaakt tussen zwakke en sterke vormen van formatief assessment. De zwakke vorm van formatief assessment is convergent, de leerkracht stelt vast of de leerling heeft geleerd, de leerkracht voert voornamelijk summatieve toetsen uit en geeft feedback op prestatie en succesvolle afronding van de taak (Cauley & McMillan, 2009). De sterke vorm van formatief assessment is divergent, door interactie tussen leerlingen en leerkrachten en leerlingen onderling wordt vastgesteld wat de leerling heeft bereikt. Deze vorm kenmerkt zich door rijke feedback die leerlingen helpt beter te leren en wordt ingezet in verschillende fasen van het leerproces waarbij de zelfregulatie en motivatie van de student worden gestimuleerd.

Op een internationale conferentie in 2009 hebben de aanwezige wetenschappelijke experts op het gebied van assessment de definitie van de sterke vorm van formatief assessment als volgt geformuleerd: *'..everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning'* (Klenowski, 2009, p. 264). Deze definitie doet volgens hen recht aan de *spirit* van formatief assessment en aan het belangrijkste doel, namelijk actieve betrokkenheid van de leerling om het leren te bevorderen (Klenowski, 2009). In onderhavig onderzoek wordt deze definitie van de sterke vorm van formatief assessment gehanteerd, omdat deze de kenmerken en componenten bevat die aanwijzingen geven hoe formatief assessment eruit kan zien en kan werken in de onderwijspraktijk van bijvoorbeeld de lerarenopleiding. De studenten staan centraal en zij verrichten samen met de lerarenopleiders en peers hun activiteiten zoals feedback geven, heldere criteria stellen en rijke vragen stellen met als doel dat de studenten autonome lerenden worden. Het is een onderzoeksproces waarbij het zoeken naar bewijs centraal staat, zodat goede beslissingen worden genomen voor de volgende stappen van studenten en lerarenopleiders. Bovendien is het een essentieel onderdeel van de dagelijkse praktijk in de klas, bedoeld om het leren te verbeteren.

Een concrete uitwerking wordt geformuleerd door Black en Wiliam (2009). Zij hebben vijf kern-elementen geformuleerd die in een klas uitgevoerd kunnen worden en bewezen effectief zijn:

1. Leerdoelen en succescriteria verduidelijken en delen;
2. Bewijs van leren verzamelen door het organiseren van effectieve (klasse-)discussie, leertaken en activiteiten te ontwerpen;
3. Feedback geven dat leren bevordert;
4. Lerenden activeren om instructiebron voor een ander te zijn;
5. Lerenden activeren om zichzelf te zien als eigenaar van hun eigen leerproces.

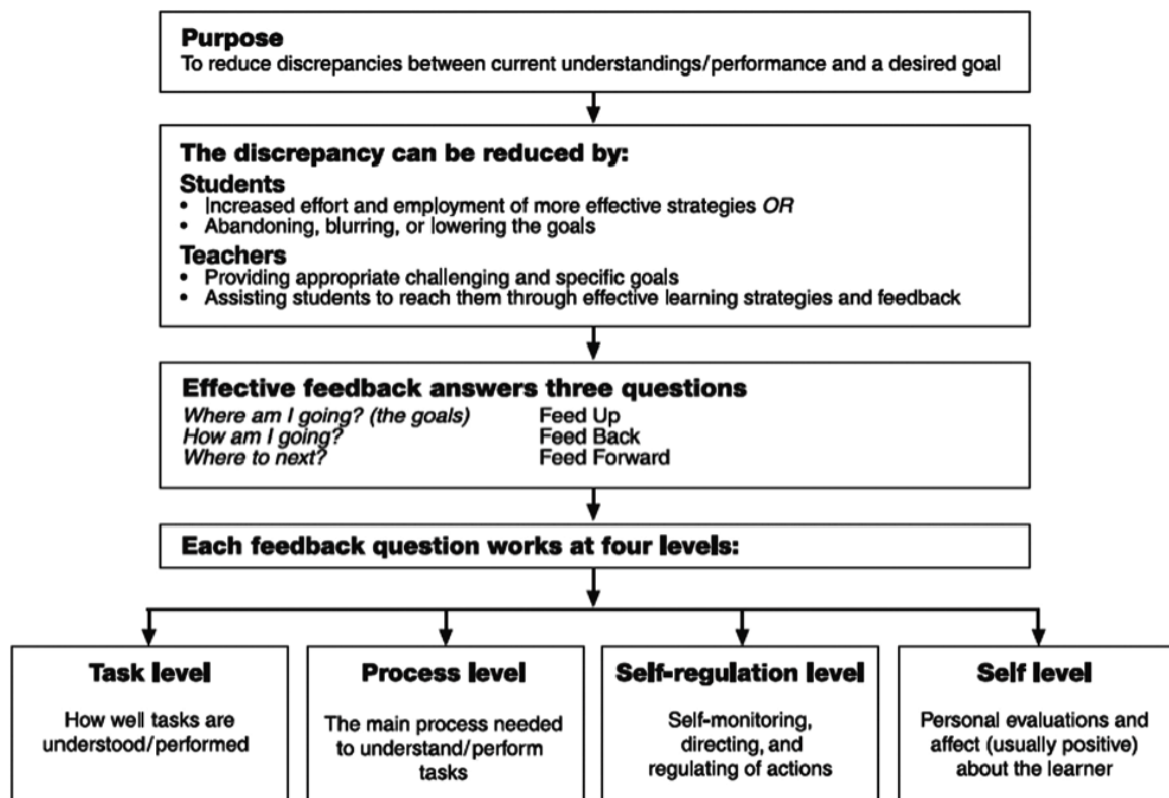
Met diverse activiteiten kunnen deze kernelementen vorm krijgen in een klaslokaal. Black en Wiliam (2009) noemen bijvoorbeeld succescriteria delen met studenten, stellen van effectieve vragen, peer-assessment, self-assessment en formatief gebruik van summatieve toetsen. Deze vijf kernelementen vormen in dit onderzoek de basis voor de operationalisering van de componenten van formatief assessment.

Bij alle activiteiten om de kernelementen vorm te geven zijn drie processen heel belangrijk. Namelijk vaststellen waar een student staat in zijn of haar leren, waar de student naar toe gaat en wat de student nodig heeft om daar te komen. De lerarenopleiders zijn hierbij verantwoordelijk voor het ontwerpen van een effectieve leeromgeving en de studenten zijn verantwoordelijk voor hun eigen leren in die omgeving. De meest besproken activiteit die bij alle kernelementen van formatief assessment een grote rol speelt, is het geven van feedback zowel door lerarenopleiders als studenten. Feedback heeft als doel het verschil tussen geleverde en gewenste prestatie te verkleinen en feedback wordt effectief als lerarenopleiders en studenten zelf weten waar ze staan en horen te zijn en welke passende activiteiten de kloof kunnen verkleinen tussen huidige resultaten en succescriteria (Sadler, 1989). Hierbij is vooral de kwaliteit en niet de kwantiteit een belangrijk aandachtspunt.

Later onderzoek van Hattie en Timperley (2007) toont aan dat effectieve feedback antwoord geeft op de vragen 'waar ga ik heen?', 'hoe sta ik ervoor?' en 'wat is de volgende stap?'. Deze vragen sluiten aan bij de drie genoemde processen en kernelementen door Black en Wiliam (2009) en refereren naar de drie categorieën waarin effectieve feedback ingedeeld kan worden: feed up, feed back en feed forward. Door feed up worden de leerdoelen, succescriteria en verwachtingsniveau helder (kernelement 1). Door feed back wordt bewijs van leren verzameld (kernelement 2) en de feed forward bestaat uit feed back dat leren bevordert (kernelement 3).

In het model van Hattie en Timperley (2007) worden deze drie feedbackvragen gekoppeld aan vier feedbackniveaus, namelijk persoonlijk, taak-, proces- en zelfregulatieniveau. Feedback op het persoonlijke niveau is gericht op kenmerken van de student in plaats van de taak, bijvoorbeeld 'je bent een goede student!'. Feedback op het taakniveau is gericht op het aangeven of de taak goed is uitgevoerd, bijvoorbeeld 'dit antwoord is onjuist'. Feedback op het procesniveau gaat over de strategieën

die nodig zijn voor de opdracht en het aangeven of er alternatieve strategieën gebruikt kunnen worden, bijvoorbeeld ‘om het juiste antwoord te vinden kun je de volgende strategie toepassen: ...’. Feedback op het zelfregulativeniveau is gericht op zelfmonitoring, evalueren en plannen. Deze feedback is bedoeld om de student te laten reflecteren op zijn of haar leerproces en dit zo nodig te verbeteren, bijvoorbeeld ‘wat zijn de overeenkomsten tussen de huidige taak en taken die je in het verleden hebt opgelost?’. In Figuur 1 wordt de relatie weergegeven tussen het doel van feedback en de manieren waarop dit bereikt kan worden.



Figuur 1. *Het feedbackmodel van Hattie en Timperley (2007) om leren te bevorderen (Hattie & Timperley, 2007, p. 87)*

Wanneer feedback gericht is op niveau van proces of zelfregulatie is feedback het meest effectief en zinvol (Hattie, 2013). De meta-analyse van Hattie (2013) laat een gemiddelde effectgrootte $d = .75$ zien voor het effectieve gebruik van feedback. Effectieve feedback is duidelijk, doelgericht, zinvol en afgestemd op voorkennis, waarbij luisteren cruciaal is in het afstemmen van feedback op de student. Feedback is duidelijk als de student de taal van de lerarenopleider begrijpt. Door feedback te relateren aan leerdoelen, gevarieerd, objectief en anekdotisch, wordt de feedback doelgericht. Feedback is zinvol wanneer deze zich op, of net boven het niveau van beheersing richt. Wanneer een student een vaardigheid goed beheerst, heeft het weinig effect om hier uitvoerig feedback op te geven.

Vanwege het belang van effectieve feedback binnen formatief assessment wordt in dit onderzoek gekozen voor de componenten feed up, feed back en feed forward op persoonlijk, taak-, proces en zelfregulatie-niveau vanuit het gezichtspunt van de lerarenopleider. In Tabel 1 worden de componenten weergegeven die de relatie vormen tussen de vijf kernelementen van Black en Wiliam (2009) en de drie feedbackvragen met de vier feedbackniveaus uit het feedbackmodel van Hattie en Timperley (2007).

Tabel 1

Componenten van formatief assessment (Black & Wiliam, 2009, p. 8)

	Feed up	Feed back	Feed forward
		persoonlijk niveau	
	taakniveau	taakniveau	taakniveau
	procesniveau	procesniveau	procesniveau
	zelfregulatie-niveau	zelfregulatie-niveau	zelfregulatie-niveau
Leraren- opleider	1. Leerdoelen en succe- scriteria verduidelijken en delen	2. Bewijs van leren verzamelen door effectieve discussie, taken en activiteiten te ontwerpen	3. Feedback geven dat leren bevordert
Peer	Leerdoelen en succescri- teria begrijpen en delen	4. Lerenden activeren om instructiebron voor een ander te zijn	
Student	Leerdoelen en succescri- teria begrijpen	5. Lerenden activeren om zichzelf te zien als eigenaar van hun eigen leerproces	

Doelstelling

Gezien de kanttekeningen en vragen bij de bestaande vragenlijsten (Gibbs & Simpson, 2003; James & Pedder, 2006; Pat-El et al., 2013) en de nadruk op de rol van effectieve feedback bij formatief assessment is het doel van dit onderzoek een samenhangende set van schalen te construeren en valideren, die inzicht geeft in studentpercepties van de mate van toepassing van feed up, feed back en feed forward op vier niveaus door de lerarenopleider, medestudenten en studenten zelf op een lerarenopleiding.

Vraagstellingen

Bovenstaande doelstelling is tweeledig en leidt tot de volgende centrale vraag in dit onderzoek: In hoeverre kan een valide en betrouwbare vragenlijst geconstrueerd worden voor het meten van studentpercepties van de mate waarin formatief assessment wordt toegepast door lerarenopleiders op een lerarenopleiding basisonderwijs en wat zijn deze studentpercepties?

Binnen dit onderzoek wordt ten aanzien van het eerste gedeelte van de centrale vraag de volgende deelvraag gesteld:

DV1: In hoeverre ondersteunt de gevonden factorstructuur het theoretische onderscheid in de verschillende aspecten van formatief assessment?

Eisen aan vragenlijst. Betrouwbaarheid en validiteit zijn sleutelbegrippen bij het beoordelen van de kwaliteit en bruikbaarheid van een meetinstrument. Betrouwbaarheid heeft betrekking op de vraag of het onderzoek door toevalsfouten verstoord wordt, of een schaal dezelfde resultaten geeft bij herhaalde toepassing onder dezelfde voorwaarden. Voor het beoordelen van de betrouwbaarheid kan er gekeken worden naar de interne consistentie, de factorstructuur, de paralleltest-betrouwbaarheid en de test-hertest-betrouwbaarheid. De laatste twee vergen een tijdsinvestering dat in dit onderzoek niet haalbaar zal zijn, de betrouwbaarheid zal daarom met Cronbach's alpha worden weergegeven (Field, 2009; Graham, Guthrie, & Thompson, 2003).

Validiteit (geldigheid) heeft betrekking op de vraag of het onderzoek door systematische fouten verstoord wordt, de mate waarin de schaal meet wat het beoogt te meten. Er wordt onderscheid gemaakt tussen verschillende soorten validiteit. Voor het beoordelen van de validiteit kan er gekeken worden naar de factoriale validiteit, de dimensionaliteit van het construct, en naar het nomologisch netwerk, de relaties tussen verschillende metingen van het construct waarbij gekeken wordt naar de subcategorieën van constructvaliditeit, de convergente en discriminante validiteit (Evers, 2010). In dit onderzoek zullen de volgende soorten validiteit worden onderzocht:

- Inhoudsvaliditeit. Met behulp van een Principale componenten analyse (PCA) en Bartlett's test of sphericity zal worden bepaald of de items homogeen zijn en slechts één construct meten (unidimensionaliteit);
- Indruksvaliditeit door beoordeling experts die evalueren of het instrument alle componenten van het concept formatief assessment meet;
- Concurrente validiteit (een variant van criteriumvaliditeit). Gegevens zouden in verschillende situaties moeten worden verzameld waarin formatief assessment op verschillende niveaus wordt uitgevoerd (van matig tot zeer goed bijvoorbeeld).
- Constructvaliditeit. Met behulp van de CFA zal worden onderzocht of en hoe goed het instrument het veronderstelde construct meet. De constructvaliditeit heeft betrekking op de theoretische inbedding van een begrip. Het is de vraag of de theoretische relaties tussen dit begrip en andere begrippen ook bestaan tussen de operationalisering van de begrippen. Het aantonen van constructvaliditeit is een lang en complex proces, waarbij vaak verschillende methoden worden gecombineerd (Evers et al., 2010; Gregory, 2007).

Ten aanzien van het tweede gedeelte van de centrale vraag worden de volgende deelvragen gesteld:
DV2: In welke mate nemen studenten de veronderstelde componenten van formatief assessment waar?

DV3: Hoe is de samenhang tussen de veronderstelde componenten van formatief assessment?

DV4: In welke mate verschillen de percepties van studenten van de mate waarin de veronderstelde componenten van formatief assessment wordt toegepast tussen de vier leerjaren?

DV5: In welke mate verschillen de percepties van studenten van de mate waarin de veronderstelde componenten van formatief assessment wordt toegepast tussen de klassen per leerjaar?

Percepties van feedback. Studenten beoordelen de wijze van toetsen en ook het ontvangen van effectieve feedback vaak negatief (Joosten-Ten Brinke, 2014). Veel toetsvormen worden namelijk gebruikt om leerprestaties te meten en bieden minimaal feedback als integraal onderdeel van het onderwijs- en leerproces (Hattie & Timperley, 2007). Een andere reden kan zijn dat de kennis en vaardigheden van toetsen en beoordelen bij lerarenopleiders niet altijd goed zijn, omdat niet alle lerarenopleiders zijn opgeleid in goede toetsing. Scholing richt zich vooral op summatief toetsen en beoordelen en nauwelijks op de formatieve functie van toetsen en beoordelen. Hieruit blijkt de verwachting dat er verschillen zullen zijn in studentpercepties van de mate waarin formatief assessment toegepast wordt door de lerarenopleiders. Deze verschillen worden ook bepaald door verschillen in epistemologische opvattingen over instructie, leren en assessment, en door verschillen in sociaal klimaat en motivatie, waarbij doelgerichtheid, leerstijl en *self-efficacy* bepalende factoren zijn (Birenbaum et al., 2011). Daarnaast worden de verschillen ook bepaald door de soort module en de mate waarin formatief assessment al een plek heeft gekregen binnen het programma van een module. Hierdoor is de verwachting is ook dat er ook verschillen zullen zijn in de studentpercepties tussen de vier leerjaren en tussen de klassen binnen een leerjaar.

Hogeschool de Kempel is geïnteresseerd in een instrument dat deze verschillen in beeld kan brengen in het kader van de professionalisering van lerarenopleiders. Enkele eisen waar een instrument met een professionaliseringsdoel aan moet voldoen, zijn:

- 'het instrument levert informatie over de richting waarin de eigen kennis en vaardigheden kunnen worden ontwikkeld;
- het instrument levert informatie over de richting waarin de eigen kennis en vaardigheden kunnen worden ontwikkeld;
- het instrument kan individueel 'just-in-time' en in een groepscontext worden ingezet' (Joosten-Ten Brinke, 2014, p. 92).

De verwachting is dat de te construeren vragenlijst een hulpmiddel kan zijn in een lerarenopleiding om het leerproces bij studenten te verbeteren en om lerarenopleiders te professionaliseren. De vragenlijst kan inzicht opleveren in de mate waarin verschillende componenten van formatief assessment worden toegepast in de perceptie van studenten. Dit inzicht zal kunnen bijdragen aan bewustwording van het eigen gedrag van de lerarenopleider en aanpassing van leren en onderwijs, waardoor meer

opbrengstgericht gewerkt kan worden en de kwaliteit van leren en beoordelen in het onderwijs daarvoor verhoogd wordt.

Methode

Participanten

Voor dit onderzoek zijn de reguliere voltijdstudenten van leerjaar 1 tot en met leerjaar 4 benaderd van Hogeschool de Kempel, een lerarenopleiding basisonderwijs. Aan het einde van het eerste kwartaal van het schooljaar 2014-2015 werd deze populatie gevraagd te participeren in dit onderzoek (bijlage A). De vragenlijst is verstuurd naar alle 666 studenten, verdeeld over 32 klassen, waarvan 90 studenten de vragenlijst niet hebben ingevuld. Deze studenten zaten in de groepen waarvoor geen tutorbijeenkomst plaats vond of waren niet aanwezig waren een tutorbijeenkomst. Hen is gevraagd de vragenlijst in eigen tijd in te vullen, respons bij deze groepen was lager dan in de groepen waarbij de onderzoeker aanwezig was. Er zijn 576 studenten gestart met de vragenlijst, waarvan 16 studenten tussentijds zijn gestopt en 5 studenten *unengaged responses* leverden. Uiteindelijk is de vragenlijst volledig en bruikbaar voor statistische verwerking voor beide delen gedeelten van de centrale onderzoeksvraag ingevuld door 555 studenten, een respons van 83,33%.

De respondenten zijn at random ingedeeld in twee groepen. De eerste groep (N = 278) is gebruikt om de onderliggende structuur van de items te onderzoeken door een Exploratieve Factor Analyse (EFA) uit te voeren. De tweede groep (N = 277) is gebruikt om de stabiliteit van de exploratieve factor structuur te toetsen door een Confirmatieve Factor Analyse (CFA) uit te voeren. De eerste groep bevatte 61 mannen (21,9%) en 217 vrouwen (78,1%), waarbij 'seks' gelijk verdeeld is over de leerjaren ($\chi^2 = 3.62$, $df = 3$, $p = .311$). De tweede groep bevatte 62 mannen (22,4%) en 215 vrouwen (76,9%), waarbij 'seks' eveneens gelijk verdeeld is over de leerjaren ($\chi^2 = 4.51$, $df = 3$, $p = .214$). In Tabel 2 wordt de vergelijking tussen de beide groepen weergegeven.

Tabel 2

Verdeling van het aantal mannen en vrouwen over de vier leerjaren voor de eerste en tweede groep

Jaar in oplei- ding	Groep 1 voor EFA				Groep 2 voor CFA			
	Aantal vrouw	Percentage in 'leer- jaar'	Aantal man	Percentage in 'leer- jaar'	Aantal vrouw	Percentage in 'leer- jaar'	Aantal man	Perce- tage in 'leer- jaar'
1	69	73,4	25	26,6	70	74,5	24	25,5
2	52	78,8	14	21,2	51	79,7	13	20,3
3	50	76,9	15	23,2	47	72,3	18	27,7
4	46	86,8	7	13,2	47	87,0	7	13

Totaal	217	78,1	61	21,9	215	76,9	62	22,4
--------	-----	------	----	------	-----	------	----	------

Procedure

Ontwikkeling vragenlijst. Bij het construeren en valideren van de vragenlijst zijn in dit onderzoek meerdere stappen gevolgd, gebaseerd op de aanbevelingen uit het onderzoek van Worthington en Whittaker (2006). De eerste stap bestond uit het selecteren en definiëren van componenten van formatief assessment. In de tweede stap werden deze items ten behoeve van de inhoudsvaliditeit voorgelegd aan lerarenopleiders en experts, inhoudsdeskundigen op het gebied van formatief assessment, met de vraag of in hun ogen de items een goede operationalisering zijn van de verschillende componenten. De derde stap was de afname van de vragenlijst. De vierde stap bestond uit statistische analyse door een EFA en een CFA uit te voeren, gevolgd door het bepalen van de interne consistenties. In de volgende paragrafen worden de verschillende stappen en gemaakte keuzes toegelicht.

Stap 1: ontwikkeling items. De formulering van de items is gebaseerd op huidige definities en operationalisering van formatief assessment in de literatuur (Bennett, 2011; Birenbaum et al., 2011; Black & Wiliam, 2009; Hattie & Timperley, 2007; Shute, 2008; Wiliam, 2011) en op items die gebruikt zijn in bestaande vragenlijsten (De Kleijn, Mainhard, Meijer, Brekelmans, & Pilot, 2013; Pat-El et al., 2013). De eerste versie van de vragenlijst bestond uit 42 items, verdeeld over drie delen (bijlage B). Het eerste deel bevatte 15 items die betrekking hadden op de activiteiten van de lerarenopleider in de groep, het tweede deel bevatte 15 items die betrekking hadden op de activiteiten van de lerarenopleider bij peer-assessment en het derde deel bevatte 12 items die betrekking hadden op de activiteiten van de lerarenopleider bij self-assessment.

Stap 2: review items. Drie experts en vier lerarenopleiders werd gevraagd de ontwikkelde items te beoordelen. De experts zijn deskundigen op het gebied van toetsen, beoordelen en leren en zijn bekend met de gebruikte literatuur. De lerarenopleiders zijn werkzaam op Hogeschool de Kempel en zijn geselecteerd op diversiteit. Een lerarenopleider natuur- en techniek, een lerarenopleider die als docent-onderzoeker deelneemt aan de kenniskring ‘Eigentijds beoordelen in het onderwijs’, de jaarcoördinator van het vierde leerjaar en de manager initieel onderwijs. Door de dialoog met deze stakeholders werd verbetering van de vorm en functie van het instrumentontwerp beoogd. De volgende vragen stonden centraal:

- Is het ontwerp relevant?
- Is de formulering van de items helder?
- Zijn de items duidelijk en leesbaar voor studenten?
- Sluit de woordkeuze in de items aan bij de terminologie die gehanteerd wordt op de opleiding?
- Zijn er aanbevelingen voor aanpassingen van onduidelijke items, toevoegen van nieuwe items of verwijderen van items?

De feedback van de stakeholders heeft geleid tot aanpassing en toevoeging van items. Vervolgens is de tweede versie door drie lerarenopleiders van Fontys Lerarenopleiding Tilburg, allen deskundig op het gebied van formatief assessment, van feedback voorzien. Hen is gevraagd om bij alle items aan te geven bij welke fase (feed up, feed back of feed forward), welke actor (docent, peer of student) en welk niveau (persoon, taak, proces of zelfregulatie) elk item behoort. De items waarover geen consensus bestond zijn aangepast. De indeling in feed up en feed forward leverde de meeste inconsistenties op. Bijvoorbeeld ‘De docent maakt duidelijk op welke manieren we het werk kunnen uitvoeren’ is aangepast naar ‘De docent maakt duidelijk op welke manieren we opdrachten kunnen gaan uitvoeren’. Uiteindelijk heeft dit proces tot de vragenlijst ‘Percepties van Feedback’ geleid, bestaande uit 65 items. Het eerste deel bevatte 27 items, het tweede en derde deel elk 19 items (bijlage C).

Stap 3: afname vragenlijst. In de voorbereiding op de definitieve afname bij de steekproef hebben vier oud-studenten van een lerarenopleiding basisonderwijs de lijst ingevuld om te bepalen hoeveel tijd het kost om deze af te nemen, dat bleek 10 tot 15 minuten te zijn. Vervolgens heeft de onderzoeker de jaarcoördinatoren en tutoren van elke klas een toelichting gegeven op het doel en de werkwijze van het onderzoek (bijlage D). De ontwikkelde vragenlijst werd online gezet met de online tool SurveyMonkey (bijlage E). Voor aanvang van de afname ontvingen de studenten een email met daarin een verzoek tot deelname aan het onderzoek en een webkoppeling naar de vragenlijst.

Bij 26 klassen vond de afname plaats in een computerruimte en zes klassen hebben de vragenlijst op een eigen mobiel of laptop ingevuld, omdat de computerruimte niet beschikbaar was. In de tutorbijeenkomst van 20 klassen heeft de onderzoeker een korte introductie gegeven op de afname, bij de overige 12 klassen hebben de tutoren zelf een korte introductie gegeven op de afname. Hierin werd benadrukt dat deelname van belang is voor kwaliteitsverbetering van het onderwijs op Hogeschool de Kempel. De afnametijd inclusief toelichting bedroeg 20 minuten. De respondenten konden de items scoren op een vijfpunts Likertschaal (1 = nooit, 2 = beperkt, 3 = regelmatig, 4 = vaak, 5 = altijd). Boven elk deel stond een introductie om de studenten te laten focussen op een door de onderzoeker bepaalde module met één lerarenopleider en om te benadrukken dat het belangrijk is om in te vullen wat de student zelf heeft ervaren, waargenomen, gezien en gehoord heeft (perceptie). Dus niet hoe de student denkt dat het zou moeten zijn (conceptie). De onderzoeker heeft de modules bepaald waarbij het criterium was dat data in verschillende situaties en op verschillende niveaus verzameld moesten worden. Voor de eerste en derdejaarsstudenten een module rekenen-wiskunde uit de integrale leerlijn en voor de tweede en vierdejaars een module onderwijskunde uit de conceptuele leerlijn, waarbij voor de tweede en vierdejaarsstudenten in het programma op de studieroute, op papier, relatief meer formatief assessment aan bod komt.

Stap 4: statistische analyse. Eerst zijn de data geanalyseerd voor de beantwoording van deelvraag 1. Vervolgens zijn de data, op basis van de gevonden factorstructuur, geanalyseerd voor de beantwoording van deelvragen 2, 3, 4 en 5.

Data-analyse

In deze paragraaf worden de uitgevoerde analyses en gehanteerde criteria beschreven die antwoord geven op de eerste deelvraag in hoeverre de gevonden factorstructuur het theoretische onderscheid ondersteunt in de verschillende aspecten van formatief assessment. Vervolgens wordt beschreven welke analyses zijn toegepast om antwoord te krijgen op de deelvragen die horen bij het tweede gedeelte van de centrale vraag.

Voorbereidende analyse. De verkregen data uit de vragenlijsten werden met SPSS versie 20 geanalyseerd. De voorbereidende analyse bestond uit het screenen van cases (*missing data*, *unengaged responses* en *outliers*) en data (*skewness* en *kurtosis*). Bij een *skewness* of *kurtosis* < -2 of > 2 werden de items niet meegenomen in de verdere analyse. Vervolgens werden de *Bartlett's test of sphericity* en *Kaiser-Meyer-Olkin's measure of sampling adequacy* uitgevoerd om te bepalen of de correlaties voldoende groot zijn voor de factor analyse en om de meest geschikte factor analyse te bepalen. Criteria hierbij waren dat bij een significante waarde van χ^2 en KMO waarde > 0.50 een Principale Component Analyse (PCA) de meest geschikte analyse is.

Exploratieve factoranalyse. Na de voorbereidende analyse volgde de PCA met SPSS versie 20. De doelen van deze analyse zijn het reduceren van data met behoud van de originele itemvariantie en het verkrijgen van een factorstructuur binnen deze steekproef met items die voldoende hoog correleren (Field, 2009; Worthington & Whittaker, 2006). Eerst werd een PCA zonder rotatie uitgevoerd om de lineaire componenten binnen de dataset te verkrijgen. Het aantal componenten met een eigenvalue > 1 en het breekpunt in de *scree plot* bepaalden hoeveel factoren verkregen konden worden. Daarna volgde een PCA met oblique rotation (promax), omdat vanuit de theorie verklaard kan worden dat de gevonden factoren met elkaar correleren. Items met lage *communalities* $< .3$ en items met overwegend lage correlaties $< .3$ of hoge correlaties $> .9$ werden verwijderd. Met behulp van de *pattern matrix* werden items die dubbel laadden en items die op een andere factoren laadden met een verschil $< .2$ verwijderd. Met behulp van de *structure matrix* werden items met een *cross-loading* $> .6$ op meerdere factoren verwijderd (Field, 2009; Graham et al., 2003). Er volgde een iteratief proces van verwijderen van items tot een eenvoudige structuur was ontstaan.

Confirmatorische factoranalyse. Vervolgens werd een CFA uitgevoerd met het lavaanpakket in R versie 3.1.2 om de gevonden factorstructuur te beoordelen en te valideren. De kwaliteit van het gevonden factormodel werd in drie stappen onderzocht door eerst na te gaan in hoeverre de door het model gereproduceerde correlaties de geobserveerde correlaties benaderden. Gehanteerde criteria hierbij waren gestandaardiseerde factorladingen $> .50$ en de residuele correlaties $< .10$. De chi-square (χ^2) en relatieve chi-square (χ^2 / df) werden buiten beschouwing gelaten, omdat deze gevoelig is voor grote steekproeven (> 250) waardoor deze bijna altijd significant is en er geen universele standaard is voor een goede fit (Kenny, 2012; Van Geel, 2012; Worthington & Whittaker, 2006). De tweede stap

bestond uit het interpreteren van de *absolute* en *incremental* model fit. *Absolute fit indices* meten hoe goed een model de gevonden relaties in de steekproef verklaart. *Incremental fit indices* meten de verbetering in een model fit door het te vergelijken met een nulmodel. De volgende indicatoren met waarden die een goede fit indiceren, werden gebruikt: (a) *Root Mean Square Error of Approximation* (RMSEA), met een waarde $< .08$ voor een goede fit en waarden tussen de $.08$ en $.10$ voor een middelmatige fit, (b) *Standardized Root Mean Square* (SRMR), met een waarde $< .08$ en absolute residuen $< .10$, (c) *Comparative Fit Index* (CFI) $> .90$ en *Tucker Lewis Index* (TLI), met een waarde $> .90$. De derde stap bestond uit het uitvoeren van modificaties bij een onvoldoende fit. Modificaties kunnen bestaan uit het uitbreiden van het model door een secundaire lading voor een item toe te voegen of door het opnemen van gecorreleerde errortermen op grond van modificatie indices (Van Geel, 2012).

Betrouwbaarheidsanalyse en beschrijvende statistiek. Van de items per component werden de interne consistenties berekend om te bepalen of de homogeniteit hoog genoeg is om de items op te tellen tot een schaal. Criterium hierbij was dat Cronbach's α rond de $.8$ moet liggen (Field, 2009). Hierna werd de vraag in welke mate studenten de verschillende componenten van feed up, feed back en feed forward percipiëren die door de lerarenopleider worden toegepast, beantwoord door de gemiddelden, standaarddeviatie en Cronbach's α te berekenen.

Correlaties en ANOVA. Vervolgens werden de correlaties tussen de gevonden factoren berekend. Deze werden bepaald met de Pearson's correlatiecoëfficiënt, waarbij de kwalificering en interpretatie gebeurde conform Dehue, Houtmans en Van Geel (2009). Om te kunnen bepalen of er significante verschillen zijn tussen de leerjaren is een variantieanalyse (ANOVA) uitgevoerd. Om tenslotte kunnen bepalen of er significante verschillen zijn tussen de klassen per leerjaar is een ANOVA uitgevoerd waarbij gebruik is gemaakt van *split file* per leerjaar (Dehue et al., 2009; Field, 2009).

Resultaten

In deze paragraaf worden de resultaten per deelvraag toegelicht.

Deelvraag 1: Betrouwbaarheid en Validiteit

Vorbereidende analyse. Na het screenen van de data bleken de items DFBp3 'De lerarenopleider geeft ons het gevoel dat we een domme klas zijn' en DFBp4 'De lerarenopleider twijfelt of we het niveau aan kunnen' niet normaal verdeeld. De *skewness* waarden van deze items waren respectievelijk 2.62 en 1.90 en de *kurtosis* waarden 6.72 en 3.04 . Bij de overige items lagen de *skewness* waarden tussen $-.53$ en $.48$ en de *kurtosis* waarden tussen $-.98$ en $.10$. De items DFBp3 en DFBp4 werden niet in de verdere analyse meegenomen. Items met overwegend lage correlaties $< .3$ of hoge correlaties $> .9$ zijn er niet. De Kaiser-Meyer-Olkin (KMO = 0.93 , 'zeer goed', ruim boven de aanvaardbare limiet van 0.5 , aldus Field, 2009) en Bartlett's test van of Sphericity ($\chi^2 (325) = 3004.40$, $p = .0$) gaven aan dat de correlaties tussen de items voldoende groot waren voor een PCA (Field, 2009). Tenslotte is geke-

ken naar gemiddelde van de *communalities*. Bij een steekproef groter dan 250 behoort deze $>.6$ te zijn en in deze steekproef ($N = 555$) bleek het gemiddelde $.612$.

Exploratieve factoranalyse. Met de overgebleven 63 items werd een PCA met oblique rotatie (promax) uitgevoerd op de eerste groep van de steekproef ($N = 278$). Omdat de items van het tweede en derde deel op één component laadden, de componenten uit grote ladingen (meer dan 11 items per component) en kleine ladingen (minder dan drie items per component) bestonden en de eerste steekproef klein is voor 63 items (ratio van aantal respondenten per item = $4.41:1$) werd de vragenlijst in drie delen gesplitst: lerarenopleider, peer-assessment en self-assessment. In de volgende subparagrafen worden de resultaten van de analyses beschreven.

EFA vragenlijst lerarenopleider. Met 23 items die betrekking hebben op de activiteiten van de lerarenopleider in de groep werd een PCA met oblique rotatie (promax) uitgevoerd. Vier componenten hadden eigenwaarden > 1 en de totaal verklaarde variantie was 58.36% . De *scree plot* was enigszins dubbelzinnig en toonde buigingen op twee en drie componenten. Het aantal componenten werd dan ook niet vastgezet in de uiteindelijke analyse. Na het iteratieve proces van verwijderen van dubbel of niet ladende items ontstond een structuur met drie componenten met een totaal verklaarde variantie van $58,65\%$. De items die clusteren op dezelfde componenten suggereren dat component 1 ‘Feedback (feed up, feed back en feed forward) op Zelfregulatie-niveau’ is, component 2 ‘Feedback (feed back en feed forward) op Taak- en Procesniveau’ en component 3 ‘Feed up op Taak- en Procesniveau’. Tabel 3 toont de factorlading na rotatie.

Tabel 3

*Overzicht van de ladingen van de items, de eigenvalue en het cumulatieve percentage verklaarde variantie per factor als resultaat van de PCA met oblique rotatie (promax) op de vragenlijst Lera-
renopleider ($N = 278$)*

Item	Lading		
	Feedback Zelfregulatie- niveau	Feedback Taak- en Procesni- veau	Feed up Taak- en Proces- niveau
DFFz De docent geeft tips hoe we in het vervolg onze planning kunnen halen	.86		
DFUz De docent geeft bij de start van de module aan hoe we de planning kunnen gaan bewaken	.84		
DFFz2 De docent vertelt ons hoe we in het vervolg onze planning beter kunnen bewaken	.83		
DFUz2 De docent geeft bij de start van de module aan hoe we ons werk kunnen gaan organiseren	.82		

DFBz2 De docent laat ons nadenken of onze planning loopt zoals deze verloopt	.66		
DFBz De docent laat ons nadenken of we met ons werk op schema liggen	.65		
DFBpr2 De docent legt uit waarom de manier waarop we ons werk hebben gemaakt geschikt is		.85	
DFBt3 De docent geeft aan wat niet goed is aan ons werk		.75	
DFBt2 De docent geeft aan wat goed is aan ons werk		.73	
DFBpr3 De docent wil weten waarom onze aanpak bij ons werk handig bleek te zijn		.63	
DFBpr4 De docent wil weten waarom onze aanpak bij ons werk niet handig is geweest		.62	
DFft2 De docent geeft aan op welke onderdelen we ons werk kunnen verbeteren		.50	
DFBt De docent stelt kritische vragen over ons eigen werk		.43	
DFUpr2 De docent geeft instructie hoe de opdrachten gemaakt kunnen gaan worden			.83
DFUpr De docent maakt duidelijk op welke manieren we opdrachten kunnen gaan uitvoeren			.75
DFUt2 De docent legt uit wat het doel is van de te maken opdrachten			.69
Eigenvalue	6.47	1.75	1.16
Cumulatief Percentage verklaarde variantie	40.44	51.41	58.65

EFA vragenlijst Peer-assessment. Met de 19 items die betrekking hebben op de activiteiten van de lerarenopleider bij peer-assessment werd een PCA met oblique rotatie (promax) uitgevoerd. Twee componenten hadden eigenwaarden > 1 en de totaal verklaarde variantie was 62.91%. De *scree plot* was enigszins dubbelzinnig en toonde buigingen op twee en vier componenten. Het aantal componenten werd dan ook niet vastgezet in de uiteindelijke analyse. Na het iteratieve proces van verwijderen van dubbel of niet ladende items ontstond een structuur met twee componenten. De items die clusteren op dezelfde componenten suggereren dat component 1 ‘Feedback op Zelfregulatie-niveau’ is en component 2 ‘Feedback op Taak- en Procesniveau’. Tabel 4 toont de factorlading na rotatie.

Tabel 4

Overzicht van de ladingen van de items, de eigenvalue en het cumulatieve percentage verklaarde variantie per factor als resultaat van de PCA met oblique rotatie (promax) op de vragenlijst Peer-assessment (N = 278)

<i>Item</i>	<i>Lading</i>
-------------	---------------

	<i>Feedback Zelfregulatie- niveau</i>	<i>Feedback Taak- en Proces- niveau</i>
De docent zorgt ervoor, dat we		
PFFz2 elkaar helpen te bedenken hoe we in het vervolg de planning beter kunnen bewaken	.92	
PFUz samen bedenken hoe we activiteiten kunnen gaan plannen	.91	
PFUz2 elkaar bij de start van de module uitleggen hoe we de planning kunnen gaan bewaken	.89	
PFBz met elkaar bekijken of we met ons werk op schema liggen	.82	
PFBz2 met elkaar evalueren of de planning loopt zoals deze verloopt	.77	
PFFz elkaar tips geven hoe we in het vervolg de planning kunnen verbeteren	.62	
PFBpr met elkaar bespreken hoe we stappen hebben toegepast bij het maken van ons werk		.91
PFBpr2 elkaar uitleggen hoe we ons werk hebben aangepakt		.87
PFFt2 elkaar uitleg geven zodat we ons werk in het vervolg beter kunnen maken		.74
PFUpr begrijpen welke stappen nodig zijn om opdrachten goed te kunnen gaan maken		.71
PFUpr2 elkaar uitleggen op welke manieren we opdrachten kunnen gaan uitvoeren		.66
PFBt3 elkaar vertellen wat we goed hebben gedaan in ons werk		.64
Eigenvalue	6.65	1.32
Cumulatief Percentage verklaarde variantie	55.40	66.41

EFA vragenlijst self-assessment. Met de 19 items die betrekking hebben op de activiteiten van de lerarenopleider bij self-assessment werd een PCA met oblique rotatie (promax) uitgevoerd. Twee componenten hadden eigenwaarden > 1 en de totaal verklaarde variantie was 61.65%. De *scree plot* bevestigde dit en toonde buigingen op twee componenten. Het aantal componenten werd dan ook vastgezet op twee componenten in de uiteindelijke analyse. Na het iteratieve proces van verwijderen van dubbel of niet ladende items ontstond een structuur met twee componenten. De items die clusteren op dezelfde componenten suggereren dat component 1 ‘Feedback op Taak- en Procesniveau’ is en component ‘Feedback op Zelfregulativeniveau’. Tabel 5 toont de factorlading na rotatie.

Tabel 5

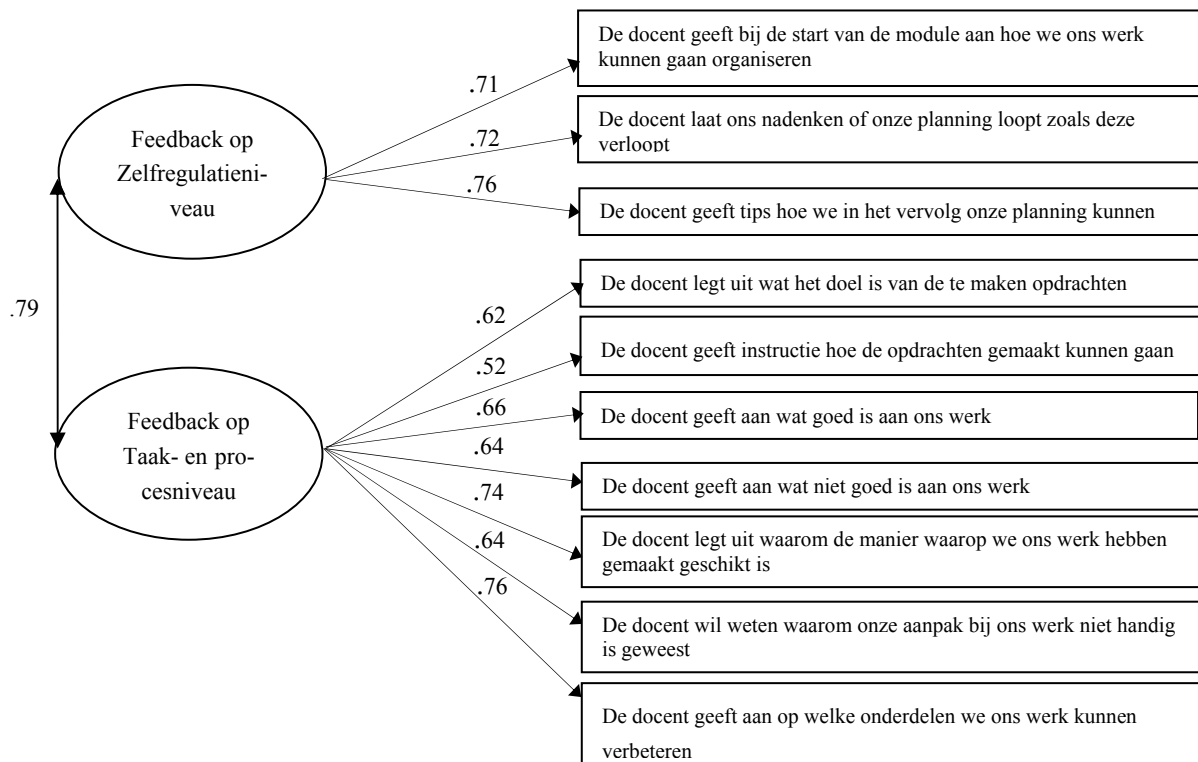
Overzicht van de ladingen van de items, de eigenvalue en het cumulatieve percentage verklaarde variantie per factor als resultaat van de PCA met oblique rotatie (promax) op de vragenlijst Self-assessment (N = 278)

<i>Item</i>	<i>Lading</i>	
	<i>Feedback Taak- en Proces- niveau</i>	<i>Feedback Zelfregulatie-niveau</i>
De docent zorgt ervoor, dat ik		
IFBt zelf weet wat ik goed heb gedaan in mijn werk	.88	
IFFpr zelf weet welke vervolgstappen ik moet zetten om mijn werk te kunnen verbeteren	.82	
IFBt2 zelf weet wat ik niet goed heb gedaan in mijn werk	.78	
IFUpr2 nadenk over de stappen die nodig zijn om de opdrachten goed te kunnen gaan maken	.77	
IFBpr zelf weet welke stappen ik heb toegepast bij het maken van mijn werk	.73	
IFUt2 zelf begrijp waarop wordt gelet bij het beoordelen van mijn werk	.72	
IFUpr weet op welke manier ik de opdrachten aan kan gaan pakken	.68	
IFFt2 weet welke onderdelen van mijn werk beter kunnen	.66	
IFBz2 zelf bekijk hoe ik mijn planning bewaak		.87
IFUz een haalbare planning voor mezelf maak voordat ik begin met studeren		.88
IFUz2 bij de start van de module weet hoe ik mijn planning kan gaan bewaken		.84
IFFz2 bedenk waar ik in het vervolg mijn planning kan verbeteren		.81
Eigenvalue	6.43	1.32
Cumulatief Percentage verklaarde variantie	53.54	64.57

Confirmatorische Factoranalyse. Met de gevonden factorstructuren werd een CFA uitgevoerd op de tweede groep van de steekproef (N = 277). De tweede groep steekproef voldeed aan het criterium 5:1 ratio van aantal respondenten per item (Worthington & Whittaker, 2006). De gevonden factorstructuren bevatten zeven manifeste variabelen ‘Feedback op zelfregulatie-niveau’ (3), ‘Feedback op Taak- en Procesniveau’ (3) en ‘Feed up Taak- en Procesniveau’ (1) met bijbehorende latente leraaropleideritems (16), peer-assessment items (12) en self-assessment items (12). In de volgende subparagrafen worden de resultaten van de analyses beschreven.

CFA vragenlijst lerarenopleider. Het drie-factorenmodel liet op basis van de grootte van de gestandaardiseerde factorladingen (allemaal $> .50$) en de residuele correlaties (15 correlaties $< .10$) vermoeden dat het model onvoldoende past bij de data. De fit indices bevestigden dit vermoeden door te laten zien dat het model van onvoldoende kwaliteit is (RMSEA = .09, SRMR = .06, CFI = .88 en TFI = .86). De modificatie indices gaven aan dat het model verbeterd kon worden door het opnemen van vijf gecorreleerde errortermen en het uitbreiden van het model met een secundaire lading voor twee items. De eerste modificatie bestond uit het toevoegen van de secundaire lading voor variabele DFUt2 en drie gecorreleerde errortermen tussen de items DFBz2 'De lerarenopleider laat ons nadenken of onze planning loopt zoals deze verloopt' en DFBz 'De lerarenopleider laat ons nadenken of we met ons werk op schema liggen', tussen DFBpr4 'De lerarenopleider wil weten waarom onze aanpak bij ons werk niet handig is geweest' en DFBpr3 'De lerarenopleider wil weten waarom onze aanpak bij ons werk handig bleek te zijn' en DFFz 'De lerarenopleider geeft tips hoe we in het vervolg onze planning kunnen halen' en DFFz2 'De lerarenopleider vertelt ons hoe we in het vervolg onze planning beter kunnen bewaken'. Deze eerste modificatie leverde een betere fit op (RMSEA = .08, SRMR = .06, CFI = .92 en TFI = .90), maar de gestandaardiseerde factorladingen op DFUt2 en DFUp2, beiden uit de factor 'Feed up op Taak- en Procesniveau', werden te laag ($< .50$).

Het opnemen van secundaire ladingen en gecorreleerde errortermen om het model te verbeteren mag niet het enige argument zijn (Van Geel, 2012). Dus een model met twee factoren werd getoetst, waarbij de laagst ladende items van de errortermen werden verwijderd. In dit model was de eerste manifeste variabele 'Feedback op Taak- en Procesniveau' met de items DFBt2, DFBpr2, DFBpr4, DFBt3, DFFt2, DFUp2 en DFUt2. De tweede manifeste variabele was 'Feedback op Zelfregulatie-niveau' met de items DFUz2, DFFz en DFBz2. Dit model leverde gestandaardiseerde factorladingen $> .50$ op en een betere fit (RMSEA = .07, SRMR = .05, CFI = .95 en TFI = .93). Figuur 2 laat het best passende model zien.

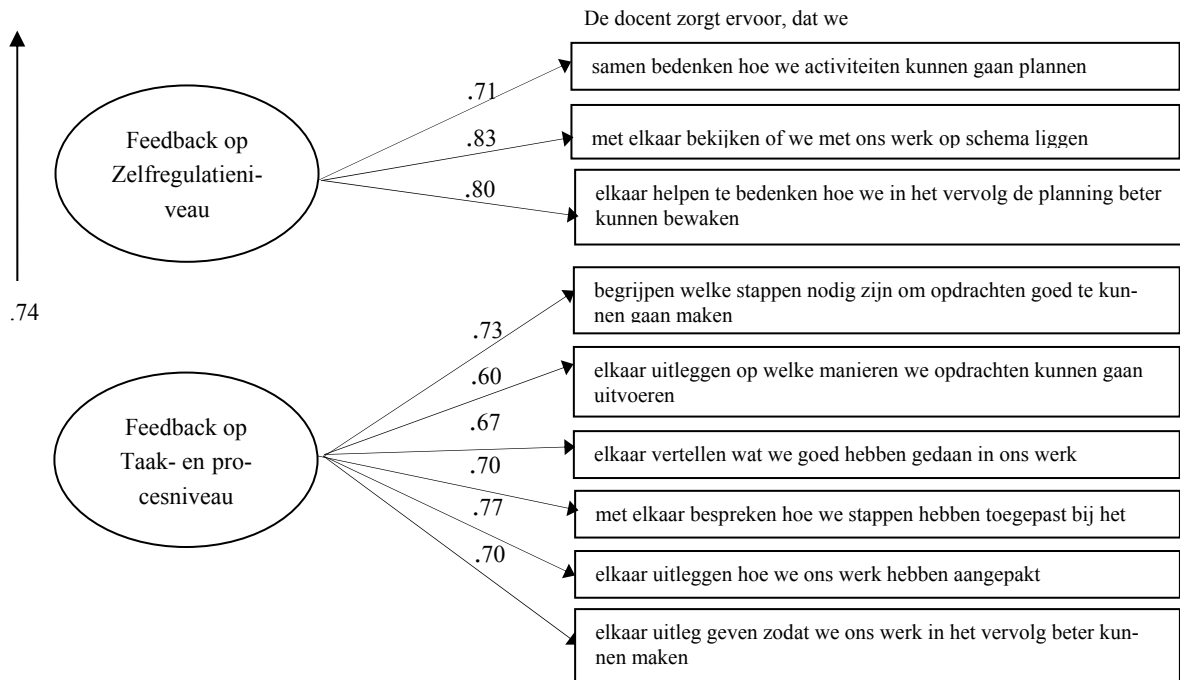


Figuur 2. De factorstructuur gebaseerd op het best passende CFA model met de gestandaardiseerde coëfficiënten op de vragenlijst Lerarenopleider

CFA vragenlijst peer-assessment. Het twee-factorenmodel liet op basis van de grootte van de gestandaardiseerde factorladingen (allemaal $> .50$) en de residuele correlaties (bijna allemaal $< .10$) vermoeden dat het model redelijk goed past bij de data. De fit indices lieten echter zien dat het model van onvoldoende kwaliteit is (RMSEA = .09, SRMR = .05, CFI = .93 en TFI = .92). De modificatie indices gaven aan dat het model verbeterd kon worden door het opnemen van de gecorreleerde errorterm tussen de items PFFz2 ‘elkaar helpen te bedenken hoe we in het vervolg de planning beter kunnen bewaken’ en PFUz ‘samen bedenken hoe we activiteiten kunnen gaan plannen’. De toepassing van deze modificatie leverde een hogere fit op (RMSEA = .08, SRMR = .04, CFI = .97 en TFI = .97). De modificatie indices in dit model gaven aan dat het model vervolgens verbeterd kon worden door het opnemen van de gecorreleerde errorterm tussen de items PFUpr ‘begrijpen welke stappen nodig zijn om opdrachten goed te kunnen gaan maken’ en PFUz ‘we samen bedenken hoe we activiteiten kunnen gaan plannen’. Deze items roepen bij respondenten niet alleen het onderliggende construct ‘feedback zelfregulatieniveau’ op, maar ook ‘planning’. De toepassing van deze modificatie leverde een hogere fit (RMSEA = .07, SRMR = .04, CFI = .96 en TFI = .94).

Ook hier gold dat het model verbeteren door het opnemen van gecorreleerde errortermen niet het enige argument mag zijn (Van Geel, 2012). Het model werd getoetst na het verwijderen van PFUz2,

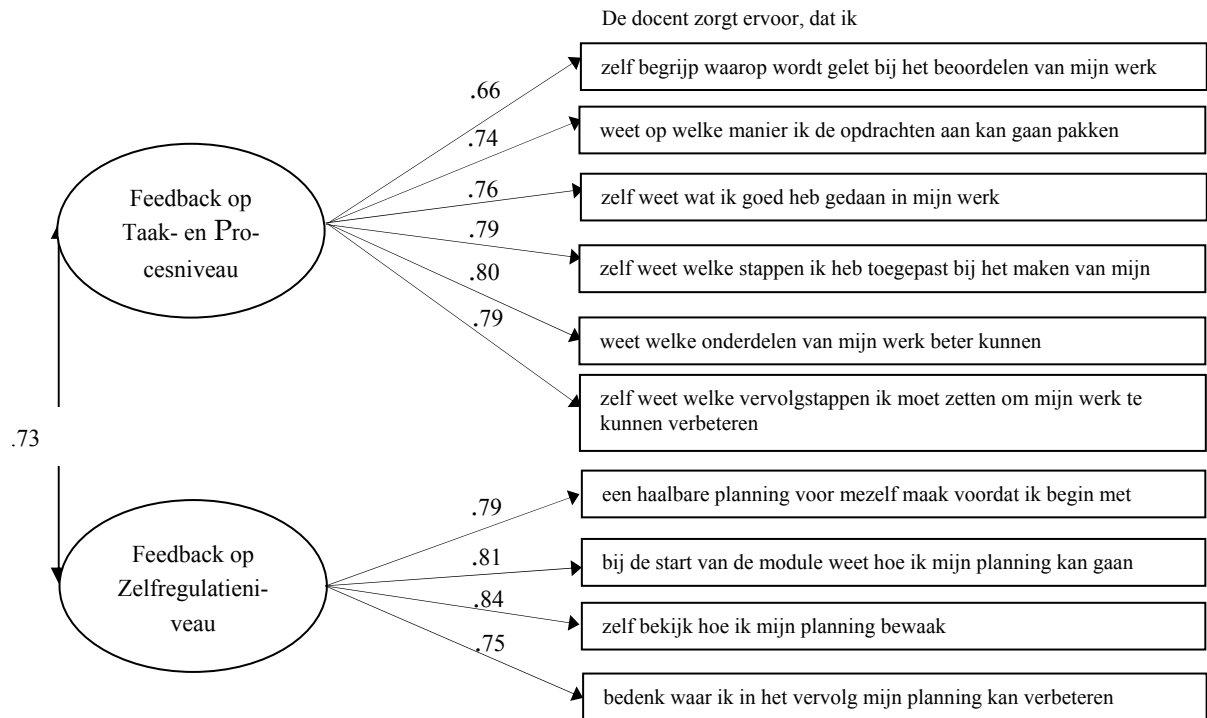
PFBz2 en PFFz, omdat er twee items over feed up, feed back en feed forward in de factor ‘Feedback op Zelfregulatienniveau’ zitten en dit waren de laagst ladende factoren. Dit model leverde een acceptabele fit op (RMSEA = .07, SRMR = .04, CFI = .97 en TFI = .96). Figuur 3 laat het best passende CFA model zien.



Figuur 3. De factorstructuur gebaseerd op het best passende CFA model met de gestandaardiseerde coëfficiënten op de vragenlijst Peer-assessment

CFA vragenlijst self-assessment. Het twee-factorenmodel liet op basis van de grootte van de gestandaardiseerde factorladingen (allemaal > .50) en de residuele correlaties (bijna allemaal < .10) vermoeden dat het model redelijk goed past bij de data. De fit indices lieten zien dat het model van voldoende kwaliteit is (RMSEA = .08, SRMR = .04, CFI = .96 en TFI = .95). De modificatie indices gaven aan dat het model verbeterd kan worden door het opnemen van de gecorreleerde errorterm tussen de items IFUpr2 ‘nadenk over de stappen die nodig zijn om de opdrachten goed te kunnen gaan maken’ en IFBpr ‘zelf weet welke stappen ik heb toegepast bij het maken van mijn werk’. Deze items vertonen inhoudelijk veel overlap. De toepassing van deze modificatie leverde een hogere fit op (RMSEA = .06, SRMR = .04, CFI = .97 en TFI = .97).

Omdat er twee items over feed up in de factor ‘Feedback op Taak- en Procesniveau’ zitten, werd het model getoetst na het verwijderen van IFUpr2, de laagst ladende factor. Dit model leverde een acceptabele fit op (RMSEA = .07, SRMR = .04, CFI = .97 en TFI = .96). Figuur 4 laat het best passende CFA model zien.



Figuur 4. De factorstructuur gebaseerd op het best passende CFA model met de gestandaardiseerde coëfficiënten op de vragenlijst Self-assessment

Betrouwbaarheidsanalyse. Na de EFA en de CFA zijn betrouwbaarheidsanalyses uitgevoerd bij de gehele steekproef. De Cronbach's α bleken voldoende hoog te zijn en de items zijn samengevoegd tot subschalen. Bij de subschaal 'Lerarenopleider Feedback op Taak- en Procesniveau' bleek uit de inter-item correlatiematrix dat DFBt3 'De lerarenopleider geeft aan wat niet goed is aan ons werk' laag correleert met DFUp2 'De lerarenopleider geeft instructie hoe de opdrachten gemaakt kunnen gaan worden' (.205). Eén van deze items verwijderen leverde geen hogere Cronbach's α op, dus deze items bleven gehandhaafd. De zes subschalen met bijbehorende items staan in bijlage F weergegeven. Vervolgens werden de gemiddelden en standaarddeviaties van de subschalen berekend. De gemiddelden van de subschalen die te maken hebben met feedback op zelfregulatieniveau lagen dicht bij elkaar ($M = 2.48, 2.54, 2.69$) en bleken onderscheidend van de subschalen die te maken hebben met feedback op taak- en procesniveau waar de gemiddelden ook dicht bij elkaar lagen ($M = 3.11, 3.12, 3.18$). Deze gegevens worden in Tabel 6 weergegeven.

Tabel 6

Beschrijvende statistiek van de subschalen (5-puntsschalen) bij de gehele steekproef ($N = 555$)

	<i>M</i>	<i>SD</i>	<i>A</i>
Peer Feedback op Zelfregulatieniveau	2.48	0.99	.88
Self Feedback op Zelfregulatieniveau	2.54	0.98	.88

Lerarenopleider Feedback op Zelfregulatie-niveau	2.69	0.86	.78
Lerarenopleider Feedback op Taak- en Procesniveau	3.11	0.69	.82
Peer Feedback op Taak- en Procesniveau	3.12	0.76	.86
Self Feedback op Taak- en Procesniveau	3.18	0.79	.88

Deelvraag 2: Mate van Studentpercepties

De gemiddelden en standaarddeviaties van de 32 klassen en elk leerjaar zijn berekend per subschaal. De beschrijvende statistieken laten zien dat de gemiddelden op de subschalen die te maken hebben met het zelfregulatie-niveau onder het rekenkundig gemiddelde 3 liggen en lager zijn dan de gemiddelden op de schalen die te maken hebben met het taak- en procesniveau die net boven het rekenkundig gemiddelde 3 liggen. De gemiddelden op alle subschalen in het eerste leerjaar zijn lager dan de gemiddelden op alle subschalen in het vierde leerjaar. In Tabel 7 worden de gemiddelden en standaarddeviaties weergegeven van de 32 klassen binnen de vier leerjaren per subschaal op zelfregulatie-niveau en in Tabel 8 op taak- en procesniveau.

Tabel 7

Weergave van de gemiddelden en standaarddeviaties van de 32 klassen binnen de vier leerjaren per subschaal op zelfregulatie-niveau bij de gehele steekproef (N = 555)

Klas/ Leerjaar	N	Lerarenopleider Feedback op Zelfre- gulatie-niveau		Peer Feedback op Zelfregulatie-niveau		Self Feedback op Zelfregulatie-niveau	
		M	SD	M	SD	M	SD
Klas 1	22	1.79	0.69	1.74	0.64	1.72	0.89
Klas 2	24	2.50	0.76	2.19	0.99	2.28	1.00
Klas 3	23	2.07	0.83	1.71	0.80	2.00	0.90
Klas 4	22	1.98	0.69	1.67	0.61	1.76	0.79
Klas 5	23	2.90	0.87	3.12	0.82	3.17	1.08
Klas 6	20	2.05	0.76	1.73	1.00	1.84	0.72
Klas 7	15	2.09	1.07	1.67	1.04	1.97	0.93
Klas 8	22	2.11	0.83	1.88	0.82	1.89	0.83
Klas 9	17	2.75	0.51	3.10	0.71	2.90	0.65
Leerjaar 1	188	2.25	0.85	2.09	0.99	2.17	1.00
Klas 10	19	3.02	0.78	2.68	0.95	2.96	0.87
Klas 11	21	2.87	0.52	2.70	0.60	2.68	0.84
Klas 12	22	2.53	0.61	2.27	0.76	2.39	1.03
Klas 13	23	2.62	0.61	2.55	0.91	2.42	0.92
Klas 14	23	2.22	0.66	2.22	0.94	2.14	0.85
Klas 15	22	2.91	0.64	2.95	0.94	2.59	0.73
Leerjaar 2	130	2.68	0.68	2.56	0.88	2.52	0.90
Klas 16	24	2.81	0.86	2.43	1.24	2.44	0.96
Klas 17	21	2.70	0.59	2.13	0.83	2.26	0.74
Klas 18	26	2.88	0.66	2.51	0.94	2.80	0.86
Klas 19	21	2.78	0.53	2.76	0.58	2.96	0.69
Klas 20	18	2.54	0.71	2.48	0.73	2.38	0.75

Klas 21	20	2.68	0.60	2.20	0.74	2.33	0.80
Leerjaar 3	130	2.74	0.67	2.42	0.89	2.54	0.84
Klas 22	9	3.30	0.45	3.26	0.70	3.61	0.38
Klas 23	11	2.97	0.66	2.94	0.80	2.77	0.59
Klas 24	8	3.38	0.72	3.21	0.67	3.03	0.75
Klas 25	7	3.00	0.92	2.67	0.98	2.75	0.87
Klas 26	13	4.15	0.48	3.56	0.76	3.83	0.49
Klas 27	10	2.73	0.80	2.17	0.95	2.03	0.62
Klas 28	9	3.11	0.60	3.04	0.93	3.17	0.75
Klas 29	9	3.59	0.85	3.48	0.84	3.53	0.88
Klas 30	12	3.75	0.70	3.42	0.49	3.65	0.84
Klas 31	13	3.77	0.53	3.44	0.84	3.40	0.86
Klas 32	6	3.22	1.05	3.39	0.98	3.00	1.11
Leerjaar 4	107	3.41	0.79	3.17	0.87	3.20	0.87
Totaal	555	2.69	0.86	2.48	0.99	2.54	0.98

Bijlage 8

Weergave van de gemiddelden en standaarddeviaties van de 32 klassen binnen de vier leerjaren per subschaal op taak- en procesniveau bij de gehele steekproef (N = 555)

		Lerarenopleider Feedback op Taak- en Procesniveau		Peer Feedback op Taak- en Procesni- veau		Self Feedback op Taak- en Procesni- veau	
Klas/ Leerjaar	N	M	SD	M	SD	M	SD
Klas 1	22	2.21	0.69	2.69	0.72	2.21	0.69
Klas 2	24	3.30	0.52	3.16	0.74	3.30	0.52
Klas 3	23	2.82	0.55	2.78	0.83	2.82	0.55
Klas 4	22	2.51	0.69	2.79	0.69	2.51	0.69
Klas 5	23	3.72	0.51	3.66	0.61	3.72	0.51
Klas 6	20	2.91	0.73	2.53	1.10	2.91	0.73
Klas 7	15	2.68	0.87	2.90	1.04	2.68	0.87
Klas 8	22	2.90	0.69	3.02	0.66	2.90	0.69
Klas 9	17	3.29	0.47	3.60	0.43	3.29	0.47
Leerjaar 1	188	2.93	0.76	3.01	0.84	2.93	0.76
Klas 10	19	3.17	0.50	3.09	0.61	3.17	0.50
Klas 11	21	3.10	0.74	2.98	0.59	3.10	0.74
Klas 12	22	3.02	0.53	2.70	0.71	3.02	0.53
Klas 13	23	3.11	0.75	2.99	0.82	3.11	0.75
Klas 14	23	2.58	0.62	2.59	0.72	2.58	0.62
Klas 15	22	3.17	0.53	3.02	0.80	3.17	0.53
Leerjaar 2	130	3.02	0.64	2.89	0.73	3.02	0.64
Klas 16	24	3.28	0.58	3.14	0.68	3.28	0.58
Klas 17	21	3.20	0.48	3.20	0.67	3.20	0.48
Klas 18	26	3.39	0.48	3.38	0.44	3.39	0.48
Klas 19	21	3.21	0.54	3.25	0.41	3.21	0.54
Klas 20	18	2.91	0.70	3.05	0.65	2.91	0.70
Klas 21	20	3.29	0.62	3.23	0.67	3.29	0.62
Leerjaar 3	130	3.23	0.57	3.22	0.59	3.23	0.57
Klas 22	9	3.29	0.48	3.61	0.38	3.29	0.48
Klas 23	11	3.03	0.51	2.77	0.59	3.03	0.51
Klas 24	8	3.43	0.53	3.03	0.75	3.43	0.53

Klas 25	7	3.16	0.88	2.75	0.87	3.16	0.88
Klas 26	13	3.66	0.63	3.83	0.49	3.66	0.63
Klas 27	10	2.90	0.61	2.03	0.62	2.90	0.61
Klas 28	9	3.32	0.50	3.17	0.75	3.32	0.50
Klas 29	9	3.79	0.59	3.53	0.88	3.79	0.59
Klas 30	12	3.57	0.55	3.65	0.84	3.57	0.55
Klas 31	13	3.43	0.51	3.40	0.86	3.43	0.51
Klas 32	6	3.57	0.83	3.00	1.11	3.57	0.83
Leerjaar 4	107	3.38	0.62	3.20	0.87	3.38	0.62
Totaal	555	3.11	0.69	3.12	0.76	3.11	0.69

Deelvraag 3: Samenhang

De correlaties tussen de zes subschalen worden in Tabel 9 weergegeven. De correlatie tussen ‘Peer Feedback op Zelfregulatie-niveau’ en ‘Self Feedback op Zelfregulatie-niveau’ duidt op een sterk positief significant verband, $r = .74$, $p < .01$ (tweezijdig getoetst). Alle overige correlaties duiden op een middelmatig positieve significante samenhang, $r = .52$ tot $.69$, $p < .01$ (tweezijdig getoetst).

Tabel 9

Correlaties tussen de zes subschalen (5-puntsschalen)

	DFLz	PFLz	IFLz	DFLtp	PFLtp	IFLtp
Lerarenopleider Feedback op Zelfregulatie-niveau	1					
Peer Feedback op Zelfregulatie-niveau	.69**	1				
Self Feedback op Zelfregulatie-niveau	.67**	.74**	1			
Lerarenopleider Feedback op Taak- en Procesniveau	.60**	.57**	.52**	1		
Peer Feedback op Taak- en Procesniveau	.54**	.63**	.54**	.64**	1	
Self Feedback op Taak- en Procesniveau	.52**	.56**	.64**	.64**	.67**	1

** Correlatie is significant bij $p < .01$, tweezijdig getoetst

Deelvraag 4: Verschillen in Studentpercepties tussen Leerjaren

De ANOVA liet zien dat de gemiddelde scores van de leerjaren per subschaal significant van elkaar verschillen. Er is een significant verschil tussen alle leerjaren op de subschalen zelfregulatie en een significant verschil tussen de leerjaren 1 en 4 op de subschalen op taak- en procesniveau. De verklaarde varianties op de schalen feedback op zelfregulatie-niveau zijn hoger dan de verklaarde varianties op de schalen feedback op taak- en procesniveau en variëren van 4,7% bij IFLtp tot 22,7% bij DFLz. In Tabel 10 worden de resultaten van de ANOVA weergegeven.

Tabel 10

Weergave van de resultaten van de verschillen tussen de vier leerjaren binnen elk leerjaar bij de gehele steekproef (N = 555)

<i>subschaal</i>	<i>toetswaarde</i>	<i>verschil</i>	<i>η²</i>
DFLz	$F(3, 551) = 53.91, p = .00$	(jaar 1) < (jaar 2 en 3) < (jaar 4)	.227
PFLz	$F(3, 551) = 31.73, p = .00$	(jaar 1) < (jaar 2 en 3) < (jaar 4)	.147
IFLz	$F(3, 551) = 29.72, p = .00$	(jaar 1) < (jaar 2 en 3) < (jaar 4)	.137
DFLtp	$F(3, 551) = 12.46, p = .00$	(jaar 1 en 2) < (jaar 2 en 3) < (jaar 3 en 4)	.064
PFLtp	$F(3, 551) = 14.37, p = .00$	(jaar 1 en 2) < (jaar 2 en 3) < (jaar 4)	.073
IFLtp	$F(3, 551) = 9.04, p = .00$	(jaar 1, 2 en 3) < (jaar 4)	.047

Voetnoot: ** $p < .05$

Deelvraag 5: Verschillen in Studentpercepties tussen Klassen binnen Leerjaar

De ANOVA liet zien dat de gemiddelde scores van de klassen binnen het eerste, tweede en vierde leerjaar op alle subschalen significant van elkaar verschillen. De gemiddelde scores van de klassen binnen het derde leerjaar verschillen niet significant van elkaar. De verklaarde varianties op de subschalen in het eerste en vierde leerjaar zijn hoger dan de verklaarde varianties op de subschalen in het tweede en derde leerjaar. In Tabel 11 worden de resultaten van de ANOVA weergegeven.

Tabel 11

Weergave van de resultaten van de verschillen tussen de klassen binnen elk leerjaar bij de gehele steekproef (N = 555)

<i>subschaal</i>	<i>toetswaarde</i>	<i>verschil</i>	<i>η²</i>
Jaar 1			
DFLz	$F(8, 179) = 4.85, p = .00$	(klas 1, 4, 6, 3, 7, 8 en 2) < (klas 4, 6, 3, 7, 8, 2 en 9) < (klas 2, 9 en 5)	.178
PFLz	$F(8, 179) = 10.43, p = .00$	(klas 4, 7, 3, 6, 1, 8 en 2) < (klas 9 en 5)	.318
IFLz	$F(8, 179) = 7.35, p = .00$	(klas 1, 4, 6, 8, 7, 3 en 2) < (klas 2 en 9) < (klas 9 en 5)	.247
DFLtp	$F(8, 179) = 11.21, p = .00$	(klas 1, 4, 7 en 3) < (klas 4, 7, 3, 8 en 6) < (klas 7, 3, 8, 6, 9 en 2) < (klas 9, 2 en 5)	.334
PFLtp	$F(8, 179) = 5.34, p = .00$	(klas 6, 1, 3, 4, 7, 8 en 2) < (klas 7, 8, 2, 9 en 5)	.193
IFLtp	$F(8, 179) = 5.15, p = .00$	(klas 1, 4, 8, 7, 6 en 3) < (klas 4, 8, 7, 6, 3, 2 en 9) < (klas 2, 9 en 5)	.187
Jaar 2			

DFLz	$F(5, 124) = 4.70, p = .001$	(klas 14, 12 en 13) < (klas 12, 13, 11, 15 en 10)	.159
PFLz	$F(5, 124) = 2.34, p = .046$	Geen significant verschil tussen de klassen binnen leerjaar 2 op subschaal PFLz.	.062
IFLz	$F(5, 124) = 2.14, p = .065$	(klas 14, 12, 13, 15 en 11) < (klas 12, 13, 15, 11 en 10)	.079
DFLtp	$F(5, 124) = 2.88, p = .017$	(klas 14, 12, 11 en 13) < (klas 12, 11, 13, 10 en 15)	.104
PFLtp	$F(5, 124) = 1.65, p = .152$	Geen significant verschil tussen de klassen binnen leerjaar 2 op subschaal PFLtp.	.062
IFLtp	$F(5, 124) = 2.34, p = .046$	(klas 14, 12, 13, 15 en 11) < (klas 12, 13, 15, 11 en 10)	.077
Jaar 3			
DFLz	$F(5, 124) = 2.68, p = .025$	Geen significant verschil tussen de klassen binnen leerjaar 3 op subschaal DFLz.	.027
PFLz	$F(5, 124) = 1.40, p = .228$	Geen significant verschil tussen de klassen binnen leerjaar 3 op subschaal PFLz.	.054
IFLz	$F(5, 124) = 2.14, p = .065$	Geen significant verschil tussen de klassen binnen leerjaar 3 op subschaal IFLz.	.098
DFLtp	$F(5, 124) = 1.65, p = .153$	Geen significant verschil tussen de klassen binnen leerjaar 3 op subschaal DFLtp.	.062
PFLtp	$F(5, 124) = 0.82, p = .535$	Geen significant verschil tussen de klassen binnen leerjaar op subschaal PFLtp.	.032
IFLtp	$F(5, 124) = 123, p = .298$	Geen significant verschil tussen de klassen binnen leerjaar 3 op subschaal IFLtp.	.047
Jaar 4			
DFLz	$F(10, 96) = 4.07, p = .000$	(klas 27, 23, 25, 28, 32, 22, 24, 29, 30 en 31) < (klas 28, 32, 22, 24, 29, 30, 31 en 26)	.298
PFLz	$F(10, 96) = 2.67, p = .006$	(klas 27, 25, 23, 28, 24, 22, 32) < (klas 25, 23, 28, 24, 22, 32, 30, 31, 29 en 26)	.218
IFLz	$F(10, 96) = 5.15, p = .000$	(klas 27, 25, 23, 32 en 24) < (klas 25, 23, 32, 24, 28, 31, 29, 22, 31 en 26)	.349
DFLtp	$F(10, 96) = 2.09, p = .032$	Geen significant verschil tussen de klassen binnen leerjaar 4 op subschaal DFLtp.	.179
PFLtp	$F(10, 96) = 4.02, p = .000$	(klas 27, 23, 25, 28 en 24) < (klas 23, 25, 28, 24, 31, 32, 22, 26, 29 en 30)	.299

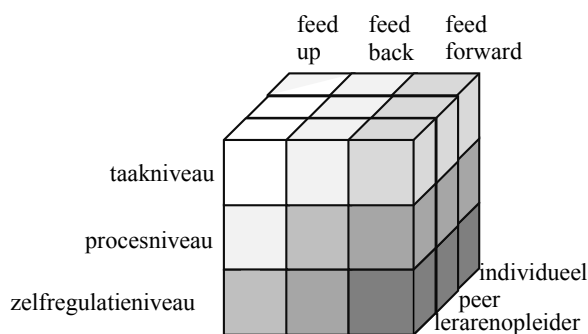
IFLtp	$F(10, 96) = 4.02, p = .000$	(klas 27, 25, 23, 28, 24, 32, 22, 29 en 31) < (klas 25, 23, 28, 24, 32, 22, 29, 31 en 30) < (klas 28, 24, 32, 22, 29, 31, 30 en 26)	.295
-------	------------------------------	---	------

Voetnoot: ** $p < .05$

Conclusie en Discussie

In de omslag van een test- naar assessmentcultuur kan de onderwijspraktijk behoefte hebben aan instrumenten die deze omslag ondersteunen en waarmee leerkrachten zich kunnen professionaliseren (Andersen et al., 2013; Robinson et al., 2014; Sluijsmans et al., 2013). Een instrument dat inzicht biedt in de percepties van studenten van de mate waarin door lerarenopleiders, peers en studenten feedback is gegeven en ontvangen zou hier aan kunnen bijdragen. De lerarenopleiders hebben daarmee een middel in handen om hun eigen onderwijs aan aankomende leerkrachten basisonderwijs te verbeteren. Er zijn echter geen valide meetinstrumenten beschikbaar om studentpercepties van formatief assessment te meten.

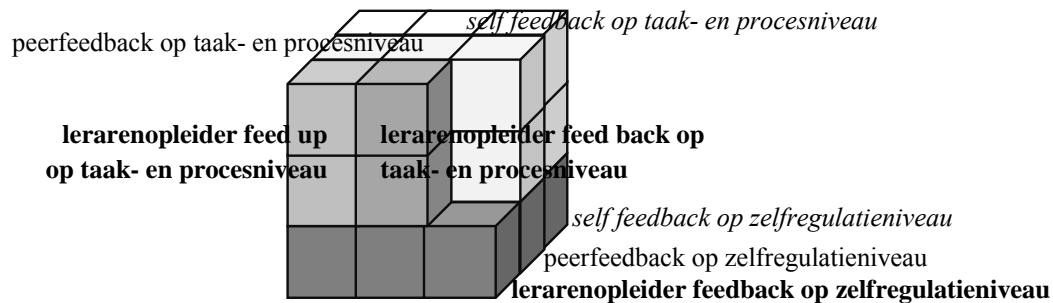
Het doel van dit onderzoek was dan ook het construeren en valideren van een samenhangende set van schalen die inzicht geeft in studentpercepties van de mate van toepassing van formatief assessment door lerarenopleiders op een lerarenopleiding. Het theoretisch model van Black en Wiliam (2009) gekoppeld aan het theoretische model van Hattie en Timperley (2007) lagen hieraan ten grondslag. Op basis van deze theorie werden componenten geoperationaliseerd. Deze waren feed up, feed back, feed forward op persoonlijk, taak-, proces- en zelfregulatie-niveau. Deze worden in Figuur 5 weergegeven.



Figuur 5. *Driedimensionale weergave van de veronderstelde theoretische negen componenten*

Uit de Exploratieve Factoranalyse bleek dat het persoonlijk niveau niet werd onderscheiden en dat de drie fasen van feedback (feed up, feed back en feed forward) en het taak- en procesniveau niet apart werden onderscheiden. Het bleek dat juist de actoren lerarenopleider, peer en individuele student apart werden onderscheiden in tegenstelling tot het veronderstelde theoretische onderscheid. Dit heeft te wellicht te maken met het opsplitsen van de vragenlijst in de drie delen lerarenopleider, peer-

assessment en self-assessment. De EFA liet een zeven-factorenstructuur zien, namelijk lerarenopleider feedback op zelfregulatie-niveau, lerarenopleider feed back op taak- en procesniveau en lerarenopleider feed up op taak- en procesniveau en zowel peer- als self-feedback op taak- en procesniveau en op zelf-regulatie-niveau. Deze worden in Figuur 6 weergegeven.



Figuur 6. Driedimensionale weergave van de zeven-factorenstructuur na EFA

De Confirmatorische Factoranalyse bevestigde dit model redelijk. Het model kon echter nog verbeterd worden door de samenvoeging van de factoren ‘lerarenopleider feed up op taak- en procesniveau’ en ‘lerarenopleider feed back op taak- en procesniveau’, waardoor uiteindelijk zes factoren werden onderscheiden. Deze worden in Figuur 7 weergegeven.



Figuur 7. Driedimensionale weergave van de factorstructuur na CFA

Vervolgens zijn de items van deze factoren bij elkaar gevoegd tot subschalen, omdat de homogeniteit van de items per subschaal voldoende hoog bleken te zijn. De resultaten lieten zien dat de gevonden subschalen voor het concept ‘formatief assessment’ betrouwbaar en valide zijn, maar wel afwijken van het vooraf veronderstelde theoretisch onderscheid. Hiermee is antwoord gegeven op de eerste deelvraag of de gevonden factorstructuur het theoretische onderscheid ondersteunt in de verschillende componenten van formatief assessment.

Een theoretische verklaring voor deze zes subschalen, waarin de fasen feed up, feed back en feed forward niet apart werden onderscheiden, kan zijn dat deze fasen teveel met elkaar correleren en niet los van elkaar gezien kunnen worden. Hattie en Timperley (2007) geven aan dat de grenzen tussen deze drie fasen vaag zijn en dat ze niet lineair geïnterpreteerd of geïmplementeerd hoeven worden.

Feedback hoeft bijvoorbeeld niet altijd te beginnen met het vragen naar de doelen (feed up), deze kunnen ook pas duidelijk worden na het vragen naar hoe we verder gaan (feed forward). In een recente studie van Arts en Jaspers (2014), waar ook de classificatie van Hattie en Timperley (2007) is gebruikt, werden deze fasen wel onderscheiden. Na analyse van feedbackopmerkingen werd echter vastgesteld dat geen feed forward en sporadisch feed up was toegepast. Dit komt enigszins overeen met de bevindingen in onderhavig onderzoek, in de subschalen zitten 12 items die feed back meten, 10 items die feed up meten en 7 items die feed forward meten (bijlage F). Hieruit rijst de vraag waarom feed forward minder wordt opgemerkt.

Een andere theoretische verklaring kan zijn dat de niveaus van feedback die door leerkrachten of peers worden geformuleerd, volgens Black en Wiliam (2009) interacteren. Feedback om begrip van een taak te verbeteren bijvoorbeeld is gerelateerd aan feedback op procesniveau om een goede aanpak te kiezen zodat de taak verbeterd wordt. Feedback op taakniveau in de vorm van goed of niet goed wordt vrijwel direct gevolgd door feedback op procesniveau in de vorm van aanwijzingen met betrekking tot de gekozen strategie of werkwijze. Dit kan verklaren waarom taak- en procesniveau niet werden onderscheiden. Feedback op zelfregulatie niveau in de vorm van vragen die aanzetten tot reflectie en interpretatie volgt hier weer op, studenten evalueren of hun taakbegrip en gebruikte strategie heeft geleid tot verbetering in hun leerproces in relatie tot hun eigen doelen en verwachtingen. Het zet aan tot autonomie, zelfcontrole, zelfsturing en zelfdiscipline (Hattie & Timperley, 2007) In dit onderzoek werd het zelfregulatie niveau wel onderscheiden van het taak- en procesniveau. Bij de items behorend bij het zelfregulatie niveau lag de nadruk op organisatie van werk, de planning van activiteiten en bewaken van de planning in zijn totaliteit.

Een methodologische verklaring voor de zes subschalen kan zijn dat het woordgebruik in de operationalisering van het zelfregulatie niveau een reden voor het clusteren van items is gebleken. Acht van de tien items bevatten het woord 'planning', waardoor andere aspecten van het zelfregulatie niveau onderbelicht bleven. De andere aspecten zijn monitoren en evalueren van het eigen leerproces, waarbij studenten actief betrokken zijn en reflecteren op hun eigen leerervaringen om de volgende stappen in hun leerproces te kunnen bepalen (Andersen et al., 2013). Hieruit rijst de vraag hoe de operationalisering verbeterd kan worden.

Een andere methodologische verklaring is te vinden in de ontwerpfase van dit onderzoek. De experts bereikten geen consensus over de indeling in fase en niveau van de verschillende items bij 17 van de 29 items in de gevonden factorstructuur. De fasen feed up en feed forward en het proces- en zelfregulatie niveau leverden de minste consensus op. Dit heeft geleid tot aanpassing van deze items in met name het werkwoordgebruik, bijvoorbeeld van 'de docent legt uit welke stappen nodig zijn om het werk goed te kunnen maken' naar 'de docent geeft instructie hoe het werk moet worden gemaakt'.

Het tweede gedeelte van de doelstelling in onderhavig onderzoek was inzicht te krijgen in de studentpercepties van de mate waarin formatief assessment wordt toegepast door lerarenopleiders op een lerarenopleiding basisonderwijs aan de hand van deelvragen twee tot en met vijf. De tweede deelvraag was in welke mate studenten de verschillende componenten van formatief assessment waarnemen. De beschrijvende statistieken laten zien dat de gemiddelden op de schalen die te maken hebben met het zelfregulatie-niveau ($M = 2.48, 2.54, 2.69$) onder het rekenkundig gemiddelde 3 liggen en lager zijn dan de gemiddelden op de schalen die te maken hebben met het taak- en procesniveau ($M = 3.11, 3.12, 3.18$) die dus net boven het rekenkundig gemiddelde liggen. Dit betekent dat de studenten regelmatig feedback op taak- en procesniveau hebben gepercipieerd en beperkt feedback op zelfregulatie-niveau. Deze bevinding sluit aan bij andere onderzoeken waar feedback op zelfregulatie-niveau ook minder of niet zijn gevonden (Arts & Jaspers, 2014; De Kleijn et al., 2013).

Verklaring hiervoor zou kunnen zijn dat volgens Hattie (2013) de niveaus corresponderen met de fases van leren, van beginnend via deskundig naar competent niveau. Dit suggereert dat feedback verschuift van taak via proces naar zelfregulatie inherent aan het beter en efficiënter beheersen van leerstof en dat feedback op het juiste moment gegeven moet worden, net op of boven het niveau van de student. Maar Hattie (2013) beschrijft ook dat studenten feedback effectiever gebruiken om het verschil te verkleinen tussen waar ze zijn in het leren en wat ze willen bereiken, als ze zichzelf monitoren en reguleren. Dat roept dan de vraag op of geconcludeerd kan worden dat studenten feedback op taak- en procesniveau meer hebben waargenomen, omdat ze een hoog niveau van zelfregulerend leren hebben. Of dat docenten minder feedback geven.

Een andere verklaring zou kunnen zijn dat het onderscheid tussen het proces- en zelfregulatie-niveau onduidelijk is. Dit bleek al bij de review van de items door drie lerarenopleiders, allen deskundig op het gebied van formatief assessment. Van de 14 initiële items op zelfregulatie-niveau werden twee items als zodanig aangegeven. De overige 12 items werden vooral aangeduid met procesniveau. In de studie van Arts en Jaspers (2014) wordt geconcludeerd dat het onderscheid tussen deze niveaus kunstmatig is, vanwege het gebruik van het woord 'strategie'. Door verschillende onderzoekers wordt het begrip 'strategie' in de context van zelfregulatie geplaatst in tegenstelling tot Hattie en Timperley (2007) die het begrip 'strategie' in de context van het procesniveau plaatsen. Dit roept de vraag op hoe het begrip 'strategie' het beste geoperationaliseerd kan worden.

De derde deelvraag was of er samenhang is tussen de verschillende kernelementen van formatief assessment. Er is een middelmatig positieve samenhang tussen alle gevonden subschalen, behalve tussen peer feedback zelfregulatie-niveau en self feedback zelfregulatie-niveau. Deze vertonen een sterk positieve samenhang: als studenten veel feedback van peers waarnemen op zelfregulatie-niveau nemen ze dat veel feedback op zelfregulatie-niveau waar bij zichzelf. De subschalen binnen hetzelfde niveau hangen meer positief met elkaar samen dan de subschalen tussen de niveaus.

De vierde deelvraag of de percepties van studenten van de mate waarin formatief assessment wordt toegepast tussen de vier leerjaren verschillen, kan positief beantwoord worden. Er bleek een significant verschil tussen de leerjaren te zijn, behalve tussen leerjaar 2 en 3 op de subschalen lerarenopleider, peer en self feedback op zelfregulatie-niveau. De eerstejaars studenten nemen feedback op zelfregulatie-niveau beperkt wat verschilt van de vierdejaars die het regelmatig waarnemen. Een verklaring kan zijn, dat perceptie te maken heeft met de fase van de opleiding waarin studenten zitten. Het eerste leerjaar bijvoorbeeld is 'selectie-georiënteerd', wat een reden kan zijn voor mindere waarneming van formatief assessment. Een andere verklaring kan zijn dat kenmerken van de context verschillen, de begeleiding van eerstejaars studenten in een reguliere Rekenen-Wiskundeles is een andere context dan de begeleiding van vierdejaarsstudenten die bezig zijn met hun 'Meesterstuk'.

De vijfde deelvraag of de percepties van studenten van de mate waarin formatief assessment wordt toegepast tussen de klassen per leerjaar verschillen, kan gedeeltelijk positief beantwoord worden. Er bleek een significant verschil tussen de klassen binnen de leerjaren 1, 2 en 4 op alle subschalen te zijn. Maar het verschil tussen de klassen binnen leerjaar 3 is niet significant. Mogelijke verklaringen zijn dat het gedrag van de docent verschilt. Binnen een leerjaar hebben verschillende docenten les gegeven.

Dus een algemene conclusie na beantwoording van deze deelvragen is dat de vragenlijst 'Percepties van Feedback' een valide en betrouwbaar instrument is dat de studentpercepties meet van de mate waarin studenten waargenomen hebben dat ze feedback op taak-, proces- of zelfregulatie-niveau hebben ontvangen van een lerarenopleider, medestudent of henzelf.

Beperkingen en Aanbevelingen

Verdere ontwikkeling, testen en valideren van de vragenlijst is aan te bevelen vanwege de theoretische en praktische relevantie voor de onderwijspraktijk. Ten eerste focust deze vragenlijst zich op feedback binnen formatief assessment, andere vragenlijsten richten zich op het gehele concept 'formatief assessment'. Deze focus is relevant, omdat effectief gebruik van feedback één van de krachtigste elementen is bij succesvol lesgeven en leren. Ten tweede is deze vragenlijst gericht op studenten in het hoger onderwijs, terwijl andere lijsten gericht zijn op andere vormen van onderwijs. Ten derde biedt deze vragenlijst de mogelijkheid voor onderzoekers om in andere contexten te onderzoeken, zoals andere modules binnen een lerarenopleiding basisonderwijs of andere lerarenopleidingen, bijvoorbeeld voortgezet onderwijs. En ten vierde bestaat de vragenlijst uit drie delen die apart en eenvoudig zijn af te nemen in groepen, waardoor het een bruikbaar instrument voor lerarenopleiders kan zijn om een beeld te krijgen van de mate waarin studenten hebben waargenomen dat ze zelf, peer of self feedback op taak- en proces niveau of zelfregulatie-niveau h

Er zijn enkele beperkingen te noemen aan dit onderzoek die aanleiding zijn voor aanbevelingen voor verder onderzoek. Ten eerste had het onderzoek een retrospectief karakter, studenten moesten zich hun percepties herinneren. Het is onduidelijk hoe betrouwbaar deze herinneringen zijn. Daarnaast

hebben studenten de neiging de feedback te herkennen die ze graag willen herkennen, ze doen aan *self-fulfilling prophecy*, waarbij vooral positieve feedback wordt geaccepteerd en negatieve feedback kritisch wordt bekeken (Hattie, 2013). Ook kan er sprake zijn van een *halo-effect*: studenten die in hun beleving goede feedback hebben ontvangen van hun lerarenopleider, gaan er als vanzelf vanuit dat die lerarenopleider bijvoorbeeld ook goede feed up of feed forward geeft. Of juist een *horn-effect*: studenten die in hun beleving slechte feedback hebben ontvangen van hun lerarenopleider, gaan er als vanzelf vanuit dat die lerarenopleider bijvoorbeeld ook goede feed up of feed forward geeft. Hierdoor rijst de vraag hoe bruikbaar de feedback van studenten is. Toekomstig onderzoek zou daarom meer prospectief kunnen zijn door percepties te monitoren tijdens een bepaalde periode. Triangulatie door de koppeling van kwantitatieve en kwalitatieve data zal de validiteit verhogen. De resultaten van dit onderzoek zouden naast een meer objectieve observatie van de feedback van de lerarenopleiders gezet kunnen worden. Bij de geselecteerde modules was geen objectieve maat beschikbaar.

Ten tweede bleek de operationalisering van feedback op zelfregulatieniveau de nadruk te leggen op planning en minder op monitoren en evalueren. Het is aan te bevelen de operationalisering verder te verbeteren. Ten derde is het onderzoek uitgevoerd op één lerarenopleiding basisonderwijs. Daarbij is gekozen voor PCA, omdat het hoofddoel reduceren van de data was. De conclusies gelden dan ook voor deze steekproef en kunnen niet zomaar gegeneraliseerd worden. Het is aan te bevelen de vragenlijst te testen op andere lerarenopleidingen om te toetsen of de gevonden factorstructuur bevestigd wordt in andere steekproeven. Ten vierde bleek de vragenlijst contextgevoelig te zijn, de percepties van studenten zijn genest in klaslokalen en antwoorden op de vragen zijn afhankelijk van bijvoorbeeld de kwaliteit van leerkrachten of het bewustzijn van doelen bij de leerlingen. Het is aan te bevelen om bij een bredere steekproef multilevel analyse uit te voeren bij een bredere steekproef, met bijvoorbeeld de niveaus leerlingen, klassen en leerjaren, waardoor minder bias ontstaat door contextgevoeligheid. Ook kan er gekeken worden naar verschillen tussen klassen over alle leerjaren, in dit onderzoek is het beperkt gebleven tot verschillen tussen klassen binnen een leerjaar.

De conclusies in dit onderzoek leiden ook tot aanbevelingen voor de praktijk. De belangrijkste vraag die in de praktijk van Hogeschool de Kempel leeft, is hoe de vragenlijst kan bijdragen aan professionalisering van de lerarenopleider. Als studenten bijvoorbeeld aangeven dat ze minder peerfeedback op taak- en procesniveau hebben waargenomen, kan dit aanleiding zijn voor de lerarenopleider te onderzoeken hoe peerfeedback vorm krijgt en hoe het verbeterd kan worden. Het vraagt echter veel van de vaardigheden bij de lerarenopleider en bij de student. Veel factoren spelen namelijk een rol bij het geven en ontvangen van effectieve feedback. Het is cruciaal een onderscheid te maken tussen *evidence-based* conclusies en interpretaties van verzamelde informatie op basis waarvan vervolgens instructie wordt aangepast. Een fout in één van beide stappen kan de effectiviteit van formatieve evaluatie verminderen (Bennett, 2011).

De betrokkenheid van lerarenopleiders bij formatief assessment zal worden verhoogd als zij zich willen professionaliseren in formatief assessment. Een iteratief trainingsproces gerelateerd aan resultaten van studenten helpt lerarenopleiders de omslag te maken in hun rol en draagt bij aan een fundamentele heroriëntatie op de relatie met de student (Robinson et al., 2014). Het toepassen van formatief assessment in de praktijk vraagt om algemene kennis van formatief assessment en bijbehorende toepassingen. Als docenten strategieën begrijpen, zullen ze deze eerder plannen en vervolgens toepassen. Belangrijke leerprincipes zijn nieuwe kennis ontwikkelen door bestaande overtuigingen erbij te betrekken, grondige kennis van feedback door nieuwe strategieën te leren en begrijpen waarom deze effectief zijn, cognitie, emotie en motivatie integreren (Timperley, 2014).

De omslag van test naar assessmentcultuur vraagt een verandering in de huidige toetspraktijk en congruentie tussen visie en uitvoering van toetsing. Deze is realiseerbaar als experts en lerarenopleiders intensief samenwerken en de denkkaders helder zijn (Hattie, 2013). De manier van denken leidt namelijk tot gewenste verandering. Het proces van vormen van nieuwe woorden en gedragingen speelt hierbij een rol. Momenteel wordt op Hogeschool de Kempel de term ‘formatief toetsen’ gebruikt. Het woord ‘toetsen’ suggereert dat formatief assessment een instrument is om het leren te bevorderen in vorm van een formatieve toets, een vragenlijst of een digitaal portfolio. Lerarenopleiders plannen een moment voor ‘formatief toetsen’, waardoor het risico bestaat dat formatief assessment geen geïntegreerd onderdeel is in de dagelijkse praktijk. Daarnaast wordt formatief tegenover summatief toetsen geplaatst, terwijl formatief gebruik van summatief toetsen een krachtig onderdeel van formatief assessment kan zijn.

‘It is the feedback information and interpretations from assessments, not the numbers or grades, that matter’ (Hattie, 2007, p.104).

Formatief assessment is juist een proces waarbij bewustwording van het eigen leren en aanwijzingen voor het leren het leerproces bevordert door ondersteuning, motivering, reflectie en dialoog in de vorm van feedback, peer-assessment en self-assessment. Proces en instrument zijn wederkerig, maar het is de conceptualisering en de kwaliteit van beide die nodig zijn. Een proces kan niet een ongeschikt instrument redden en een instrument kan niet een ongeschikt proces vervangen (Bennett, 2011).

Referenties

- Andersen, I., Castelijns, J., & Van Ravenswaaij, H. (2013). *Formatieve assessment in de praktijk. Onderzoeksrapportage*. 's-Hertogenbosch: KPC Groep in opdracht van het ministerie van OCW.
- Arts, J., & Jaspers, M. (2014). Kwaliteit van schriftelijke feedback. In D. Joosten-Ten Brinke, J. Arts, C. Burghout, M. Jaspers, W. Lansu, M. Maas, J. Verhoosel & S. Kratsborn (Eds.), *Eigentijds toetsen en beoordelen. De opbrengst van vier jaar praktijkonderzoek* (pp. 11-20). Tilburg: Fontys Lerarenopleiding Tilburg.
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18 (1), 5-25.
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., . . . Nickmans, G. (2006). A learning Integrated Assessment System. *Educational Research Review*, 1 (1), 61-67.
- Birenbaum, M., Kimron, H., & Shilton, H. (2011). Nested contexts that shape assessment "for" learning: School-based professional learning community and classroom culture. *Studies in Educational Evaluation*, 37 (1), 35-48.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21 (1), 5-31.
- Blok, H., Ledoux, G., & Roeleveld, J. (2013). Opbrengstgericht werken in het primair onderwijs: theorie en praktijk. Retrieved 12 februari, 2014, from http://www.kohnstamminstituut.uva.nl/pdf_documenten/Artikel%20OPBRENGSTGERICHT%20WERKEN%20online%20versie_271113.pdf
- Broers, S. (2014). De Ontwikkeling van een curriculum Eigentijds Beoordelen in het Onderwijs. Eindrapportage. Helmond: Lectoraat Eigentijds Beoordelen in het Hoger Onderwijs, Hogeschool de Kempel.
- Castelijns, J., Heldens, W., Wijnen, M., & Wouters, M. (2011). Eigentijds Beoordelen in het Onderwijs. Deelrapport 1: Bouwstenen. Helmond: Lectoraat Eigentijds Beoordelen in het Onderwijs, Hogeschool De Kempel.
- Cauley, K. M., & McMillan, J. H. (2009). Formative assessment techniques to support student motivation and achievement. *Clearing House*, 83 (1), 1-6.
- De Kleijn, R. A. M., Mainhard, M. T., Meijer, P. C., Brekelmans, M., & Pilot, A. (2013). Master's thesis projects: Student perceptions of supervisor feedback. *Assessment & Evaluation in Higher Education*, 38 (8), 1012-1026.
- Dehue, F., Houtmans, T., & Van Geel, R. (2009). *Onderzoekspracticum kwantitatieve data-analyse*. Heerlen: Open Universiteit Nederland.
- Dochy, F., Schelfhout, W., & Janssens, S. (2003). *Anders evalueren. Assessment in de onderwijspraktijk*. Leuven: LannooCampus.
- Driscoll, M. P. (2005). *Psychology of learning for instruction (3rd ed.)*. Boston: Allyn and Bacon.

- Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment. *Higher Education*, 22 (3), 201-204.
- Evers, A. K. W. R. R. (2010). The Dutch Review Process for Evaluating the Quality of Psychological Tests: History, Procedure, and Results. *International Journal of Testing*, 10 (4), 295-317.
- Field, A. (2009). *Discovering statistics using IBM SPSS statistics. Third edition*. Londen: Sage.
- Gibbs, G., & Simpson, C. (2003). Measuring the response of students to assessment: the Assessment Experience Questionnaire. *Improving student learning. Theory and practice*, 10.
- Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling*, 10 (1), 142-153.
- Haroldson, R. A. (2012). *Student Perceptions of Formative Assessment in the Chemistry Classroom (Doctoral dissertation)*. Retrieved from <http://conservancy.umn.edu>
- Hattie, J. (2013). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77 (1), 81-112.
- James, M., & Pedder, D. (2006). Beyond method: assessment and learning practices and values. *Curriculum Journal*, 17 (2), 109-138.
- Johnson, M., & Burdett, N. (2010). Intention, interpretation and implementation: Some paradoxes of assessment for learning across educational contexts. *Research in Comparative and International Education*, 5 (2), 122-130.
- Joosten-Ten Brinke, D. (2014). Bewust en bekwaam toetsen en beoordelen. In D. Joosten-Ten Brinke, J. Arts, C. Burghout, M. Jaspers, W. Lansu, M. Maas, J. Verhoosel & S. Kratsborn (Eds.), *Eigentijds toetsen en beoordelen. De opbrengst van vier jaar praktijkonderzoek*. Tilburg: Fontys Lerarenopleiding Tilburg.
- Kenny, D. A. (2012). *Measuring Model Fit*. Retrieved 23 december, 2014, from <http://davidakenny.net/cm/fit.htm>.
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16 (3), 263-268.
- Könings, K. D., Brand-Gruwel, S., & Van Merriënboer, J. G. (2005). Towards more powerful learning environments through combining the perspectives of designers, teachers, and students. *British Journal of Educational Psychology*, 75 (4), 645-660.
- Korver, B., & Tillema, H. (2014). Feedback provision in mentoring conversation - differing mentor and student perceptions. *Journal Of Education And Training Studies*, 2 (2), 167-175.
- Lizzio, A., & Wilson, K. (2004). First-year students' perceptions of capability. *Studies in Higher Education*, 29 (1), 109-128.

- Loughran, J. (2012). *What expert teachers do: Enhancing professional knowledge for classroom practice*: Routledge.
- Ministerie van OC&W. (2011). Actieplan 'Basis voor Presteren': Naar een ambitieuze leercultuur voor alle leerlingen. Retrieved 12 februari, 2014, from <http://www.rijksoverheid.nl/actieplannenocw>
- Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2013). Validation of Assessment for Learning Questionnaires for teachers and students. *British Journal of Educational Psychology*, 83 (1), 98-113.
- Perumanathan, P. S. (2014). Formative assessment and feedback in the primary classroom: An interplay between teachers' beliefs and practices (Doctoral thesis). Retrieved from <http://researcharchive.vuw.ac.nz>
- Robinson, J., Myran, S., Strauss, R., & Reed, W. (2014). The impact of an alternative professional development model on teacher practices in formative assessment and student learning. *Teacher Development*, 18 (2), 141-162.
- Rusman, E., Boon, J., Martínez-Monés, A., Rodríguez-Triana, M. J., & Retalis, S. (2013). Towards the Use of New Methods for Formative e-Assessment of 21st Century Skills in Schools.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18 (2), 119-144.
- Schulz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback: USA-Colombia. *The Modern Language Journal*, 85 (2), 244-258.
- Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation*, 37 (1), 49-54.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78 (1), 153-189.
- Sluismans, D. M. A., Joosten-ten Brinke, D., & Van der Vleuten, C. P. M. (2013). *Toetsen met leerwaarde. Een reviewstudie naar de effectieve kenmerken van formatief toetsen*. Den Haag: Nederlandse Organisatie voor Wetenschappelijk Onderzoek.
- Struyven, K., Dochy, F., & Janssens, S. (2003). Students' perceptions about new modes of assessment in higher education: A review. In M. Segers, Dochy, F., & Cascallar, E. (Ed.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 171-223): Springer Netherlands.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30 (4), 325-341.

- Struyven, K., Dochy, F., Janssens, S., Schelfhout, W., & Gielen, S. (2006). De effecten van de evaluatie-ervaring op de percepties van studenten over assessment. *Pedagogische Studiën*, 83 (1), 35.
- Valcke, M. (2010). *Onderwijskunde als ontwerpwetenschap: een inleiding voor ontwikkelaars van instructie en voor toekomstige leerkrachten*. Gent: Academina Press.
- Van den Berg, D., & Vandenberghe, R. (1999). *Succesvol leiding geven aan onderwijsinnovaties. Investeren in mensen*. Alphen aan den Rijn: Samson.
- Van der Kleij, F., Vermeulen, J., Eggen, T., & Veldkamp, B. (2013). Leren van toetsen; een cyclisch proces. Retrieved 2 september, 2014, from http://toetswijzer.kennisnet.nl/html/leren_van_toetsen/default.shtm
- Van Geel, R. (2012). *Confirmatorische Factoranalyse met Lavaan*. Heerlen: Open Universiteit Nederland.
- Vermeulen, J. A., & Van der Kleij, F. M. (2012). Towards an integrative formative approach of data-driven decision making, assessment for learning, and diagnostic testing.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37 (1), 3-14.
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research a content analysis and recommendations for best practices. *The Counseling Psychologist*, 34 (6), 806-838.

Bijlage A Verzoek tot deelname aan onderzoek

Beste student,

De Kempel wil zich blijven verbeteren en heeft daar jouw hulp bij nodig!

Dit onderzoek richt zich op het gebruik van feedback door lerarenopleiders en studenten in het onderwijs op de Kempel.

Aan alle reguliere voltijdstudenten wordt gevraagd deze vragenlijst in te vullen over de mate waarin zij feedback hebben gekregen. Ik wil ook jou vragen deze lijst in te vullen. Het invullen van deze vragenlijsten zal ongeveer 15 minuten in beslag nemen.

Alle gegevens die met deze vragenlijsten worden verzameld zullen anoniem blijven. De gegevens zullen vertrouwelijk worden behandeld en worden alleen voor dit onderzoek gebruikt.

In de vragenlijst staan uitspraken. Lees iedere uitspraak goed en sla geen uitspraak over. Nadat je de vragenlijst geopend hebt, wordt per onderdeel precies toegelicht wat er van je verwacht wordt.

Wanneer je na afloop een samenvatting van het onderzoek wenst te ontvangen, kun je aan het einde van de vragenlijst je e-mailadres achterlaten.

Ik dank je bij voorbaat hartelijk voor je deelname en bijdrage aan het onderzoek met het invullen van deze vragenlijst.

Beste student,

De Kempel wil zich blijven verbeteren en heeft daar jouw hulp bij nodig!

Dat kun je doen door vandaag in de TUB* een vragenlijst in te vullen over het gebruik van feedback door lerarenopleiders en studenten in het onderwijs op de Kempel.

In overleg met je tutor is er in de TUB van vandaag tijd ingeroosterd om deze vragenlijst in te vullen. Deze kom ik toelichten en dan gaan we naar ICT-1, zodat je op de computer de vragenlijst in kunt vullen. Dit is de link naar de vragenlijst:

<https://nl.surveymonkey.com/s/VragenlijstK1-Kempel>

Ik dank je bij voorbaat hartelijk voor je deelname en bijdrage aan het onderzoek met het invullen van deze vragenlijst.

Sandra Broers

student Onderwijswetenschappen Open Universiteit

*TUB = Tutorbijeenkomst

Bijlage B Eerste versie van de vragenlijst

Feed up	Docent	Peer	Student
Doelen les	De docent verduidelijkt de lesdoelen	De docent stimuleert ons de lesdoelen aan elkaar uit te leggen	De docent stimuleert ons zelf na te denken over de lesdoelen
Doelen module	De docent verduidelijkt de toetsdoelen van. (I-OND5, MST, RW, NED)	De docent stimuleert ons de toetsdoelen van ... (I-OND5, MST, RW, NED) aan elkaar uit te leggen	De docent stimuleert ons zelf na te denken over de toetsdoelen van ... (I-OND5, MST, RW, NED)
Beoordelingscriteria	De docent legt uit waarop wordt gelet bij de beoordeling van de toets	De docent stimuleert ons aan elkaar uit te leggen waarop wordt gelet bij de beoordeling van de toets	De docent stimuleert ons dat we begrijpen waarop wordt gelet bij de beoordeling van de toets
Proces	De docent bespreekt met ons de stappen die nodig zijn om de toets goed te kunnen maken	De docent stimuleert ons aan elkaar uit te leggen welke stappen nodig zijn om de toets goed te kunnen maken	De docent stimuleert ons zelf na te denken over de stappen die nodig zijn om de toets goed te kunnen maken
Zelfregulatie	De docent bespreekt met ons de globale planning in deze module	De docent stimuleert ons de globale planning in deze module aan elkaar uit te leggen	De docent stimuleert ons zelf na te denken over de globale planning in deze module

Feedback	Docent	Peer	Student
Leertaak om bewijs van begrip te verzamelen	De docent praat met ons over hoe het gaat met onze opdrachten	De docent stimuleert ons elkaar te praten over hoe het gaat met onze opdrachten	De docent stimuleert ons om zelf na te denken over hoe het gaat met onze opdrachten
Leertaak om bewijs van begrip te verzamelen	De docent stelt kritische vragen over ons eigen werk	De docent stimuleert ons kritisch naar elkaars werk te kijken	De docent stimuleert ons om zelf kritisch naar ons eigen werk te kijken
Taak	De docent vertelt ons wat we goed en niet	De docent stimuleert ons elkaar te vertellen wat we	De docent stimuleert ons om zelf na te denken over

	goed doen bij onze opdrachten	goed en niet goed doen bij onze opdrachten	wat goed en niet goed gaat bij onze opdrachten
Proces	De docent vraagt ons hoe we ons werk aanpakken	De docent stimuleert ons elkaar te vertellen hoe we ons werk aanpakken	De docent stimuleert ons om zelf na te denken over de manier waarop we ons werk aanpakken
Zelfregulatie	De docent controleert of we met ons werk op schema liggen	De docent stimuleert ons om elkaar te controleren of wij met ons werk op schema liggen	De docent stimuleert ons om zelf onze planning te bewaken

Feed forward	Docent	Peer	Student
Leren bevorderen	De docent geeft extra uitleg bij het maken van ons werk als we iets nog niet begrijpen	De docent stimuleert ons elkaar te helpen als we iets nog niet begrijpen	De docent stimuleert ons om zelf aan te geven dat we iets nog niet begrijpen
Taak	De docent vertelt ons op welke onderdelen we ons werk kunnen verbeteren	De docent stimuleert ons elkaar te vertellen welke onderdelen van ons werk beter kunnen	De docent stimuleert ons om zelf na te denken welke onderdelen van ons werk beter kunnen
Proces	De docent legt ons uit met welke stappen we ons werk kunnen verbeteren	De docent stimuleert ons elkaar uit te leggen met welke stappen we ons werk kunnen verbeteren	De docent stimuleert ons om zelf na te denken met welke stappen we ons werk kunnen verbeteren
Zelfregulatie	De docent geeft tips hoe we onze planning kunnen halen	De docent stimuleert ons elkaar tips te geven voor onze planning	De docent stimuleert ons zelf de planning te verbeteren

Bijlage C Items van de vragenlijst 'Percepties van Feedback'

Feed up

DFUt	De docent bespreekt de criteria die gebruikt worden bij de beoordeling van ons werk
DFUt2	De docent legt uit wat het doel is van de te maken opdrachten
DFUt3	De docent legt uit wat het doel is van een les
DFUt4	De docent legt uit wat de doelen zijn van de module
DFUpr	De docent maakt duidelijk op welke manieren we opdrachten kunnen gaan uitvoeren
DFUpr2	De docent geeft instructie hoe de opdrachten gemaakt kunnen gaan worden
DFUz	De docent geeft bij de start van de module aan hoe we de planning kunnen gaan bewaken
DFUz2	De docent geeft bij de start van de module aan hoe we ons werk kunnen gaan organiseren

Feedback

DFBp	De docent vertelt ons dat we een slimme klas zijn
DFBp2	De docent laat merken dat we het niveau aan kunnen
DFBp3	De docent geeft ons het gevoel dat we een domme klas zijn
DFBp4	De docent twijfelt of we het niveau aan kunnen
DFBt	De docent stelt kritische vragen over ons eigen werk
DFBt2	De docent geeft aan wat goed is aan ons werk
DFBt3	De docent geeft aan wat niet goed is aan ons werk
DFBpr	De docent legt uit waarom de manier waarop we ons werk gemaakt hebben geschikt is
DFBpr2	De docent legt uit waarom de manier waarop we ons werk hebben gemaakt niet geschikt is
DFBpr3	De docent wil weten waarom onze aanpak bij ons werk handig bleek te zijn
DFBpr4	De docent wil weten waarom onze aanpak bij ons werk niet handig is geweest
DFBz	De docent laat ons nadenken of we met ons werk op schema liggen
DFBz2	De docent laat ons nadenken of onze planning loopt zoals deze verloopt

Feed-

Forward

DFFt	De docent geeft aan hoe in het vervolg de uitwerking eruit zou moeten zien
DFFt2	De docent geeft aan op welke onderdelen we ons werk kunnen verbeteren
DFFpr	De docent legt ons uit met welke vervolgstappen we ons werk kunnen verbeteren
DFFpr2	De docent bespreekt met ons wat we nog moeten weten om ons werk goed te kunnen maken
DFFz	De docent geeft tips hoe we in het vervolg onze planning kunnen halen
DFFz2	De docent vertelt ons hoe we in het vervolg onze planning beter kunnen bewaken

Feed up

- PFUt we met elkaar de doelen bespreken voordat we de opdrachten gaan maken
- PFUt2 we elkaar uitleggen waarop gelet zal worden bij de beoordeling van ons werk
- PFUpr we begrijpen welke stappen nodig zijn om opdrachten goed te kunnen gaan maken
- PFUpr2 we elkaar uitleggen op welke manieren we opdrachten kunnen gaan uitvoeren
- PFUz we samen bedenken hoe we activiteiten kunnen gaan plannen
- PFUz2 we elkaar bij de start van de module uitleggen hoe we de planning kunnen gaan bewaken

Feed-back

- PFBt we elkaar kritische vragen stellen over ons gemaakte werk
- PFBt2 we elkaar vertellen wat we goed hebben gedaan in ons werk
- PFBt3 we elkaar vertellen wat we goed hebben gedaan in ons werk
- PFBpr we met elkaar bespreken hoe we stappen hebben toegepast bij het maken van ons werk
- PFBpr2 we elkaar uitleggen hoe we ons werk hebben aangepakt
- PFBz we met elkaar bekijken of we met ons werk op schema liggen
- PFBz2 we met elkaar evalueren of de planning loopt zoals deze verloopt

Feed-**Forward**

- PFFt we elkaar duidelijk maken welke onderdelen van ons werk in het vervolg beter kunnen
- PFFt2 we elkaar uitleg geven zodat we ons werk in het vervolg beter kunnen maken
- PFFpr we elkaar uitleggen met welke vervolgstappen we ons werk kunnen verbeteren
- PFFpr2 we elkaar informeren wat we moeten doen om ons werk een volgende keer beter te kunnen maken
- PFFz we elkaar tips geven hoe we in het vervolg de planning kunnen verbeteren
- PFFz2 we elkaar helpen te bedenken hoe we in het vervolg de planning beter kunnen bewaken

Feed up

- IFUt zelf nadenk over de doelen die ik moet halen
- IFUt2 zelf begrijp waarop wordt gelet bij het beoordelen van mijn werk
- IFUpr weet op welke manier ik de opdrachten aan kan gaan pakken
- IFUpr2 nadenk over de stappen die nodig zijn om de opdrachten goed te kunnen gaan maken
- IFUz een haalbare planning voor mezelf maak voordat ik begin met studeren
- IFUz2 bij de start van de module weet hoe ik mijn planning kan gaan bewaken

Feed-**back**

- IFBt ik zelf weet wat ik goed heb gedaan in mijn werk
-

IFBt2	ik zelf weet wat ik niet goed heb gedaan in mijn werk
IFBpr	ik zelf weet welke stappen ik heb toegepast bij het maken van mijn werk
IFBpr2	ik zelf evalueer of mijn aanpak bruikbaar is geweest bij het maken van mijn werk
IFBpr3	ik zelf evalueer waarom mijn aanpak niet bruikbaar is geweest bij het maken van mijn werk
IFBz	ik systematisch mijn eigen werk observeer aan de hand van beoordelingscriteria
IFBz2	ik zelf bekijk hoe ik mijn planning bewaak

Feed-

Forward

IFFt	mijn aandacht richt op de onderdelen die moeilijk gaan
IFFt2	weet welke onderdelen van mijn werk beter kunnen
IFFpr	zelf weet welke vervolgstappen ik moet zetten om mijn werk te kunnen verbeteren
IFFpr2	weet welke informatie ik nodig heb om mijn werk te verbeteren door de feedback die ik heb ontvangen, weet hoe ik in het vervolg mijn planning beter kan
IFFz	bewaken
IFFz2	bedenk waar ik in het vervolg mijn planning kan verbeteren

Bijlage D Toelichting voor tutoeren op het doel en de werkwijze van het onderzoek

Beste tutor,

De afgelopen weken ben ik bezig geweest met het ontwikkelen van een vragenlijst. Deze vragenlijst meet de mate waarin feedback wordt toegepast in het onderwijs op de Kempel. Om deze vragenlijst te kunnen valideren is een grote steekproef noodzakelijk. De vragenlijst wordt aan alle reguliere voltijd-studenten voor gelegd. Aan hen wordt gevraagd in welke mate zij feed up, feedback en feed forward hebben waargenomen tijdens de lessen en bijeenkomsten van één module in blok 1. Het gaat om de tutorbijeenkomsten van MST-14, de lessen van C-RW9 voor de K2-studenten, de lessen en bijeenkomsten van I-OND5 voor de K1-studenten en de lessen van C-RW1 voor de P-studenten.

Ik wil je medewerking vragen bij het afnemen van de vragenlijst.

In de week van 27 oktober tot en met 31 oktober hebben de jaarcoördinatoren een kwartier tot half uur gereserveerd in de TUB voor het afnemen van de vragenlijst. De vragenlijst wordt digitaal afgenomen, daarom heb ik de ICT-ruimtes gereserveerd om met je groep daar de vragenlijst in te kunnen vullen. Ik ben er zelf dan ook, dus ik kan de afname nog toelichten en assisteren bij eventuele vragen.

Op de dag van de TUB ontvangen jij en de studenten een link naar de vragenlijst. Nadat ze op een computer in de ICT-ruimte zijn ingelogd, klikken ze de link aan en dan wijst de rest zich vanzelf.

Alvast bedankt voor je medewerking.

Met vriendelijke groeten,

Sandra Broers

Bijlage E Vragenlijst 'Percepties van Feedback'

Perceptie van Feedback
<p>Beste student,</p> <p>De Kempel wil zich blijven verbeteren en heeft daar jouw hulp bij nodig! Dit onderzoek richt zich op het gebruik van feedback door docenten en studenten in het onderwijs op de Kempel.</p> <p>Aan alle reguliere voltijdstudenten wordt gevraagd deze vragenlijst in te vullen over de mate waarin zij feedback hebben gekregen. Ik wil ook jou vragen deze lijst in te vullen. Het invullen van deze vragenlijst zal ongeveer 15 minuten in beslag nemen.</p> <p>Alle gegevens die met deze vragenlijst worden verzameld zullen anoniem blijven. De gegevens zullen vertrouwelijk worden behandeld en worden alleen voor dit onderzoek gebruikt.</p> <p>In de vragenlijst staan uitspraken. Lees iedere uitspraak goed en sla geen uitspraak over. Nadat je de vragenlijst geopend hebt, wordt per onderdeel precies toegelicht wat er van je verwacht wordt.</p> <p>Wanneer je na afloop een samenvatting van het onderzoek wenst te ontvangen, kun je aan het einde van de vragenlijst je e-mailadres achterlaten.</p> <p>Ik dank je bij voorbaat hartelijk voor je deelname en bijdrage aan het onderzoek met het invullen van deze vragenlijst.</p> <p>Sandra Broers student Onderwijswetenschappen Open Universiteit</p>
Perceptie van Feedback
<p>Algemene informatie</p> <p>Wat is je leeftijd?</p> <p><input type="text"/></p> <p>In welke klas zit je?</p> <p><input type="text"/></p> <p>Wat is je geslacht?</p> <p><input type="radio"/> man</p> <p><input type="radio"/> vrouw</p> <p>Deze vragenlijst gaat over de mate waarin jij hebt waargenomen dat jouw docent feedback heeft gegeven tijdens de lessen van één module in blok 1.</p> <p>Over welke module ga je de vragen beantwoorden?</p> <p><input type="text"/></p> <p>Welke docent heb je in die module gehad?</p> <p><input type="text"/></p>

Feedback in de klas

Deze vragenlijst gaat over de mate waarin jij hebt waargenomen dat jouw docent onderwijskunde feedback heeft gegeven tijdens de lessen van I-OND5. Hieronder volgen 27 uitspraken. Lees deze goed en geef voor elke uitspraak aan in hoeverre jij dit hebt waargenomen tijdens de lessen en bijeenkomsten van I-OND5 in blok 1. Je kunt kiezen uit:

- ☐ niet
- ☐ beperkt
- ☐ regelmatig
- ☐ vaak
- ☐ altijd

Het is belangrijk om in te vullen wat jijzelf ervaren, waargenomen, gezien, gehoord hebt. Niet over hoe je denkt dat het zou moeten zijn.

	niet	beperkt	regelmatig	vaak	altijd
De docent legt uit wat de doelen zijn van de module	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent vertelt ons hoe we in het vervolg onze planning beter kunnen bewaken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent legt uit wat het doel is van de te maken opdrachten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent geeft bij de start van de module aan hoe we ons werk kunnen gaan organiseren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent geeft tips hoe we in het vervolg onze planning kunnen halen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent legt uit waarom de manier waarop we ons werk gemaakt hebben geschikt is	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent legt uit wat het doel is van een les	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent vertelt ons dat we een slimme klas zijn	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent wil weten waarom onze aanpak bij ons werk handig bleek te zijn	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent geeft aan op welke onderdelen we ons werk kunnen verbeteren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent geeft bij de start van de module aan hoe we de planning kunnen gaan bewaken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent bespreekt met ons wat we nog moeten weten om ons werk goed te kunnen maken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De docent wil weten waarom onze aanpak bij ons werk niet handig is geweest	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Perceptie van Feedback

Beste student,

hartelijk dank voor het invullen van deze vragenlijst.

Als je nog opmerkingen of vragen hebt, kun je die hieronder aangeven.

Je kunt je emailadres achterlaten als je een samenvatting van het onderzoeksrapport wilt ontvangen.

Sandra Broers

Ruimte voor opmerkingen en/of vragen

Vul hieronder je emailadres in (optioneel).

Bijlage F Items van de zes gevalideerde subschalen

Leraren- opleider	Feed up	Feedback	Feed forward
Taak- en Proces- niveau	De docent legt uit wat het doel is van de te maken opdrachten	De docent geeft aan wat goed is aan ons werk	De docent geeft aan op welke onderdelen we ons werk kunnen verbeteren
	De docent geeft instructie hoe de opdrachten gemaakt kunnen worden	De docent geeft aan wat niet goed is aan ons werk	
		De docent legt uit waarom de manier waarop we ons werk hebben gemaakt niet geschikt is	
		De docent wil weten waarom onze aanpak bij ons werk niet handig is geweest	
Zelf- Regulatie- niveau	De docent geeft bij de start van de module aan hoe we ons werk kunnen gaan organiseren	De docent laat ons nadenken of onze planning loopt zoals deze verloopt	De docent geeft tips hoe we in het vervolg onze planning kunnen halen
Peer	Feed up	Feedback	Feed forward
Taak- en Proces- niveau	Als we in een groepje werken, zorgt de docent ervoor dat:	Als we in een groepje werken, zorgt de docent ervoor dat:	Als we in een groepje werken, zorgt de docent ervoor dat:
	we begrijpen welke stappen nodig zijn om opdrachten goed te kunnen gaan maken	we elkaar vertellen wat we goed hebben gedaan in ons werk	we elkaar uitleg geven zodat we ons werk in het vervolg beter kunnen maken
	we elkaar uitleggen op welke manieren we opdrachten kunnen gaan uitvoeren	we met elkaar bespreken hoe we stappen hebben toegepast bij het maken van ons werk	

		we elkaar uitleggen hoe we ons werk hebben aangepakt	
Zelf-Regulatie-niveau	we samen bedenken hoe we activiteiten kunnen gaan plannen	we met elkaar bekijken of we met ons werk op schema liggen	we elkaar helpen te bedenken hoe we in het vervolg de planning beter kunnen bewaken
Student	Feed up	Feedback	Feed forward
	Als ik individueel werk, zorgt de docent ervoor dat ik:	Als ik individueel werk, zorgt de docent ervoor dat ik:	Als ik individueel werk, zorgt de docent ervoor dat ik:
Taak-en Proces-niveau	zelf begrijp waarop wordt gelet bij het beoordelen van mijn werk	ik zelf weet wat ik goed heb gedaan in mijn werk	weet welke onderdelen van mijn werk beter kunnen
	weet op welke manier ik de opdrachten aan kan gaan pakken	ik zelf weet welke stappen ik heb toegepast bij het maken van mijn werk	zelf weet welke vervolgstappen ik moet zetten om mijn werk te kunnen verbeteren
Zelf-Regulatie-niveau	een haalbare planning voor mezelf maak voordat ik begin met studeren	ik zelf bekijk hoe ik mijn planning bewaak	bedenk waar ik in het vervolg mijn planning kan verbeteren
	bij de start van de module weet hoe ik mijn planning kan gaan bewaken		